Cranfield University

Daniel Trujillano Lidón

The role of non-coding RNAs in haemoglobin regulation

Cranfield Health

Master of Science by Research

Cranfield University

Cranfield Health

Master of Science by Research

2009

Daniel Trujillano Lidón

The role of non-coding RNAs in haemoglobin regulation

Supervisor: Dr David Carter

Academic Year 2008 to 2009

ABSTRACT

Non-coding RNAs appear to play a role in gene regulation by modulating chromatin structure. There is mounting evidence suggesting an essential role for non-coding RNAs in the complex process of the genetic regulation of the β-globin locus. Preliminary observations indicate that the BGL3 non-coding transcript may be involved in an RNA-protein interaction and may be interacting with chromatin in the β-globin locus as part of a regulatory function within the locus. However, the expression profile of this non-coding transcript has not yet been characterized and nothing is known about its mode of action.

Here it is shown that the BGL3 transcript is dynamically up-regulated upon haemin induction of the K562 cell line (a human erythroleukemic cell line). To determine whether there is a correlation between the BGL3 transcript expression and the expression of the γ- and β-globin genes, the levels of the BGL3 transcript in K562 cells were perturbed by knocking it down using the RNA interference pathway. The effect of the knockdown of the BGL3 transcript was tested on the expression levels of the γ- and β-globin genes, which were quantified using qRT-PCR. Our results are the first, to our knowledge, that describe a developmentally regulated expression of the BGL3 non-coding transcript in haemin-induced K562 cells, and provide evidence that suggests that this transcript may be involved in the silencing of the β-globin gene.

Keywords:

BGL3 transcript, human β-globin locus, intergenic transcription, K562 cells, globin switching.

ACKNOWLEDGEMENTS

First of all, I would like to thank Dr David Carter for giving me the opportunity to undertake this project, and for his guidance throughout all the MRes.

I would like to express my deepest gratitude to my fiancée Alba. Without her love and patience this thesis would not have been possible.

I would also like to thank to my family for all his support during all my university studies.

Finally, I would like to thank Dan Shreve and Ryan Pink for all the fun we have had in the lab and during the meetings.

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

This literature review aims to explore the current pool of knowledge of the molecular mechanisms that regulate the production of red blood cells from precursor stem cells, as well as the mechanisms that control the structure of DNA at the genes that code for adult and foetal haemoglobins involved in haemoglobin switching. The molecular and cellular mechanisms of red blood cell development will be discussed, with special emphasis on how expression of different haemoglobin subunits is controlled, the molecular mechanism of red blood cell differentiation, and how non-coding RNAs can regulate both of these processes. A better understanding of the molecular and cellular mechanisms of globin switching is a prerequisite in the development of rational therapies for genetic blood disorders such as sickle cell disease or thalassemias.

1.1 The β-globin Locus

1.1.1 Haematopoietic system

The formation of blood cellular components through the differentiation of hematopoietic stem cells is known as haematopoiesis. Haematopoiesis is carried out by the haematopoietic system, a well–characterized vertebrate developmental system with a highly regenerative capability which maintains tissue homeostasis. It controls commitment, proliferation, apoptosis, and maturation of hematopoietic stem and progenitor cells (HSCs) in response to environmental signals. These environmental changes are perceived by the HSCs, which can undergo self-renewal division or differentiation to produce cells of all the blood lineages through a regulated and rapid mechanism[1]. This is a sustainable process that continues throughout the lifespan of an individual, responsive to temporal physiological needs, such as red cell replacement after acute blood loss or leukocytosis in response to infection. Decades of research have been devoted to the elucidation of the transcriptional networks involved in haematopoiesis. However, it remains unclear how this system dynamically adjusts to the body's needs by replenishing specific lineages of mature blood cells from uncommitted HSCs. The control of haematopoiesis and its output in terms of cell types and quantities, rely on a complex network of transcription and growth factors[2].

Haematopoiesis takes place in different anatomic sites during development. In the embryonic stage it occurs in the yolk sac, liver, and in later stages in the bone marrow. In adults, haematopoiesis takes place in marrow and lymphatic tissues. These changes of site during development are governed by cell-fate diversification and differentiation[1]. Large-scale gene-expression profiling studies of uncommitted HSCs and intermediate progenitor cells showed promiscuity in the expression of genes before lineage commitment and differentiation, as megakaryocitopoiesis, macrophagopoiesis, erythropoiesis and granulocytopoiesis share a common differentiation pathway[2]. This promiscuity of HSCs suggests that an important stage of during lineage differentiaition may be the down-regulation of groups of genes[3-5].

HSCs are the common precursor of the distinct cell lineages of mammalian blood cells. HSCs must pass through many critical decision points to differentiate into mature blood cells. The first step consists of the differentiation of HSCs to common myeloid progenitors (CMPs) or to common lymphoid progenitors (CLPs). Subsequently, CMPs differentiate into erythrocytes, megakaryocytes, myeloblasts or mast cells, whilst CLPs differentiate into T or B cells or natural killer cells. The hematopoietic system (Figure 1-1) provides a good model to study cell lineage specification because a large amount of information is available about its cellular and transcriptional networks[1].

The formation of normal mature blood cells requires the haematopoietic differentiation to be carried out in an organized spatial and temporal manner. Haematopoiesis is a complicated developmental programme mediated by a large number of regulators. Once the progenitors have chosen a specific cell fate, they maintain the lineage specification and cannot differentiate to other lineages because of the inactivation of the genes of these other lineages[2].

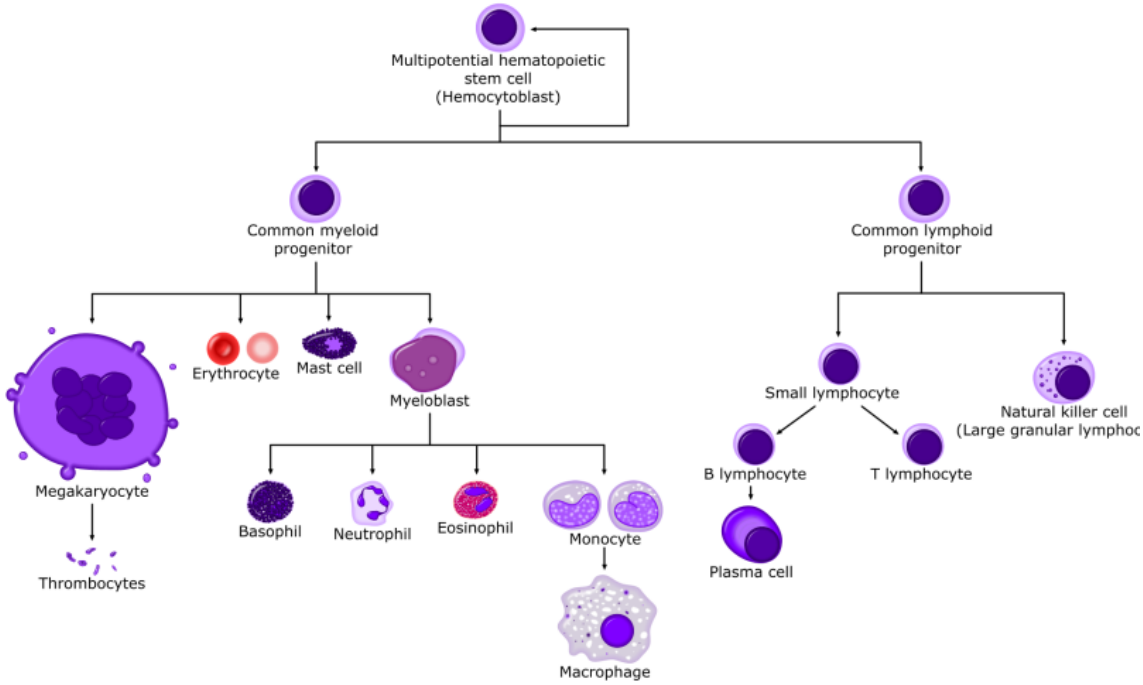


Figure 1- 1 The haematopoietic system.

Diagrammatic representation of the bone marrow pluripotent stem cell and the cell lines that arise from it. Source: Hoffbrand, 2005[2].

1.1.2 Erythropoieisis

The formation of red blood cells is known as erythropoiesis. It is a dynamic and sequential process which occurs in different anatomic sites during development. Erythroid differentiation begins with the commitment of hematopoietic stem cells (HSCs) to the erythroid lineage (Figure 1-2). Erythroid maturation consists of a series of cell transformations from erythroid progenitors cells colony-forming unit granulocyte, erythroid, monocyte and megakaryocyte (CFU-GEMM), erythroid colony-forming units (CFU-Es) and erythroid burst-forming units (BFU-Es), to morphologically distinct erythroid precursors called pronormoblasts, basophilic erythroblasts, polychromatic erythroblasts and orthochromatophilic erythroblasts, sequentially. Reticulocytes appear once the orthochromatophilic erythroblasts extrude the nucleus. As reticulocytes lose their polyribosomes they become mature red blood cells[1].

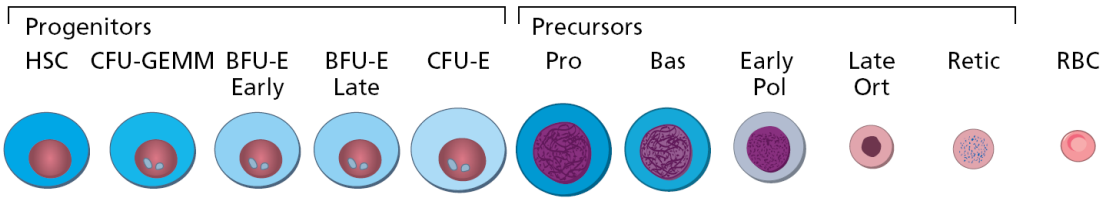


Figure 1- 2 Erythropoietic differentiation.
Pronormoblasts (Pro), basophilic erythroblasts (Bas), polychromatic erythroblasts (Pol), orthochromatic erythroblasts (Ort), reticulocytes (retic), mature red blood cells (RBCs). Adapted from: Hoffbrand, 2005, p.15[2].

Morphology, gene expression, as well as growth factor dependence are different in erythroid cells at different developmental and differentiation stages. For example, as erythroid terminal differentiation progresses, cells gradually decrease in volume, there are increases in haemoglobinization and chromatin condensation, erythropoietin signalling becomes essential and the cell-surface erythroid-specific Ter119 antigen is expressed[2].

1.1.3 Haemoglobin

Eyrthrocytes carry O_2 to the tissues and CO_2 from the tissues to the lungs, using haemoglobin[6]. Given the different oxygen requirements during the embryonic, foetal and adult stage, different haemoglobins are synthesized at each of these developmental stages. All types of haemoglobin consist of two different pairs of globin chains, each attached to one haem molecule, to achieve a final tetrameric structure. Two possible configurations are available for haemoglobin; oxygenated and deoxygenated (Figure 1-3)[2].

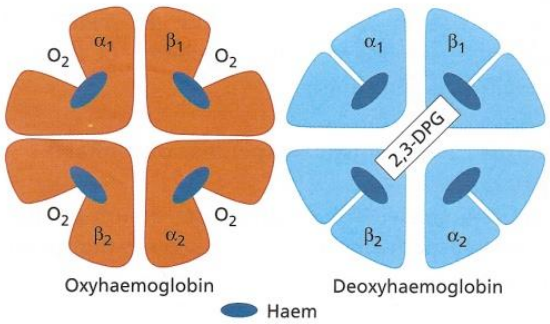


Figure 1- 3 The two possible conformations of the haemoglobin molecule.
α- and β-globin chains of normal adult haemoglobin (HbA). 2,3-DPG, 2,3-diphosphoglycerate. Source: Hoffbrand, 2006, p.17[1].

In vertebrate red cells, different haemoglobin tetramers are expressed and assembled at different developmental stages in a programmed manner[7]. In embryos, ζ-chains combine with γ-chains to form Portland ($\zeta_2\gamma_2$) haemoglobin, or with ε-chains to generate Gower 1 ($\zeta_2\varepsilon_2$) haemoglobin. When α- and ε-chains combine, Gower 2 ($\alpha_2\varepsilon_2$) haemoglobin is produced. During foetal life foetal haemoglobin is expressed (HbF, $\alpha_2\gamma_2$), whereas in adult life the major haemoglobin is haemoglobin A (HbA, $\alpha_2\beta_2$). However, during adult life haemoglobin A_2 (HbA_2, $\alpha_2\delta_2$) is also expressed together with HbF, but both represent less than 2% of the total haemoglobin[8] (Figure 1-4). A mutation un the TATA-box region of the δ-globin gene determines its low expression levels [9]. Human foetal β-like γ-globin genes and the production of HbF in primates represent a rather recent event in evolution respect to other mammals; for example mice don't have γ-globin genes. The development of these genes is thought to be due to the increased oxygen affinity of HbF over HbA, which enhances an improved oxygen delivery to the foetus in the placental circulation[8; 10].

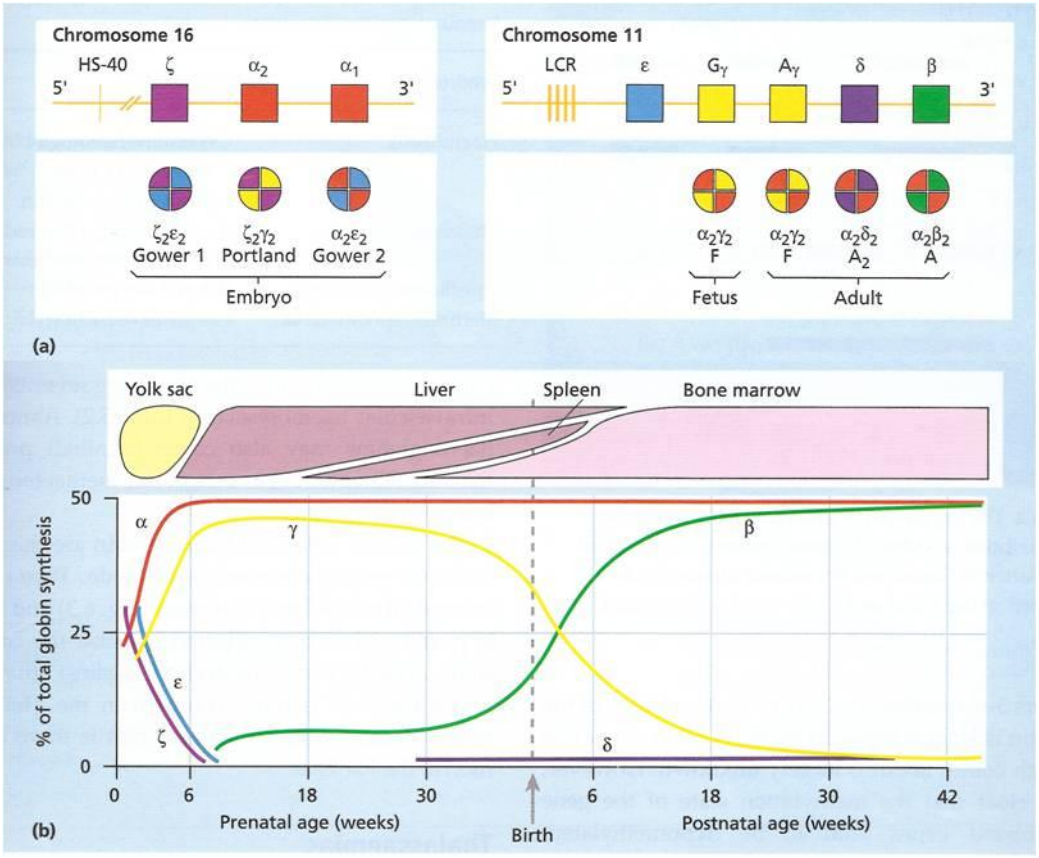


Figure 1- 4 (a) The α- and β- globin gene clusters on chromosomes 16 and 11 respectively. In embryonic, foetal and adult life different globin genes are activated or suppressed. The different globin chains are synthesized independently and then combine with each other to produce the different haemoglobins. **(b) Synthesis of individual globin chains in prenatal and postnatal life.** Adapted from: Hoffbrand, 2006, p.73[1].

In humans the genes of both α- and β-globin clusters are arranged in the order of their expression, reflecting the sequence of their expression during development[7; 9]. The β-like globin genes form a linked cluster on chromosome 11 and are arranged in the order 5′-ε-^{G}γ-^{A}γ-ψβ-δ-β-3′. The α-like globin genes also form a linked cluster, located on chromosome 16, and are arranged in the order 5′-ζ-ψζ-ψα$_1$-α$_2$-α$_1$-3′. The ψβ, ψζ and ψα pseudogenes sequences are similar to those of the β-, ζ- or α-genes, but present mutations that determine its inviability to translated to proteins. It is thought that these pseudogenes could have been functional at previous stages of evolution[2], or non-functional duplications of coding genes, or they might be active regulators of gene

expression, since endogenous siRNAs (reviewed in section 1.3) have been found in mouse embryos that come from pseudogenes[11].

1.1.4 β-globin locus

The human β-globin locus spans over 70 kb in an AT-rich region on chromosome 11 embedded within an array of olfactory receptor genes[12; 13] and contains the five β-globin genes (ε, $^{G}\gamma$, $^{A}\gamma$, δ and β), which are arranged in the order of their expression during development[8; 14; 15] (Figure 1-5). Unlike the α-globin locus which is present in a GC-rich domain on chromosome 16 and whose chromatin structure is constitutively accessible, the β-globin locus shows a tissue-specific expression limited to the cells of the erythroid lineage[16]. This is supported by the fact that the locus is DNase I-sensitive exclusively in erythroid cells[13].

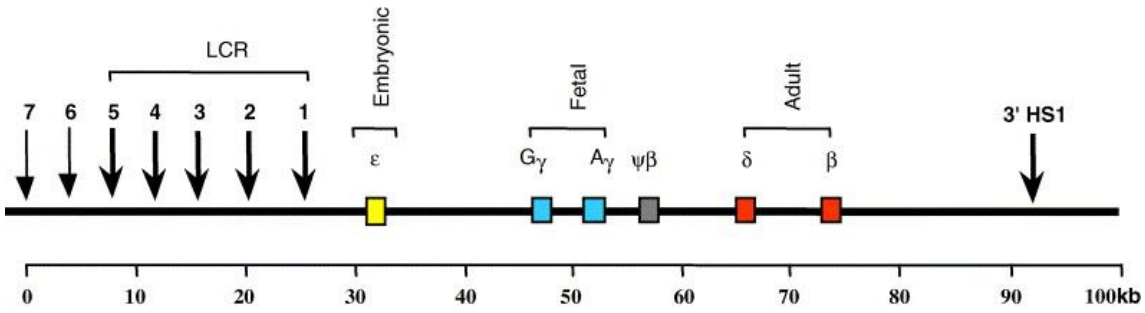


Figure 1- 5 Diagram of the β-globin locus.

A linear map with the globin locus control region (LCR) and its hypersensitive (HS) sites is indicated by the vertical arrows. The structural ε-, $^{G}\gamma$-, $^{A}\gamma$-, δ- and β- globin genes as well as the ψβ gene. Source: Stamatoyannopoulos, 2005,p.263[10].

1.1.5 Haemoglobin switching

In mammals there is a correlation between the β-globin genes expressed in each of the discrete stages of erythropoiesis and the changing sites of erythropoiesis during development[10]. Humans undergo two major transitions in the β-globin locus expression in accordance to the oxygen requirements of the foetus during its different developmental stages[6; 12]. The first gene of the β-globin locus to be expressed during the early embryonic stage is the β-like ε-globin gene in the erythroid cells located in the blood islands of the foetal yolk sac, where the primitive human embryonic

erythropoiesis first occurs. This is followed by the activation of the major β-like γ-genes ($^{G}\gamma$, $^{A}\gamma$) during foetal development, when erythropoiesis takes place in the liver and $^{G}\gamma$- and $^{A}\gamma$-globin genes are transcribed in the definitive circulating enucleated erythrocytes. Finally the β-like δ-globin and the β-globin genes are activated in the late foetal stage in the erythroid cells in the bone marrow where erythropoiesis occurs in adults, being the β-globin gene the most highly expressed in adults[10; 17] (Figure 1-6). Thus, each gene in the β-globin locus is differentially transcribed during development. This tissue-specific and developmental regulation is achieved thanks to the participation of both gene proximal and gene distal elements[18; 19]. Although these globin gene switches represent one of the classic examples of developmental control in eukaryotes, little is known about the mechanisms that mediate them[18].

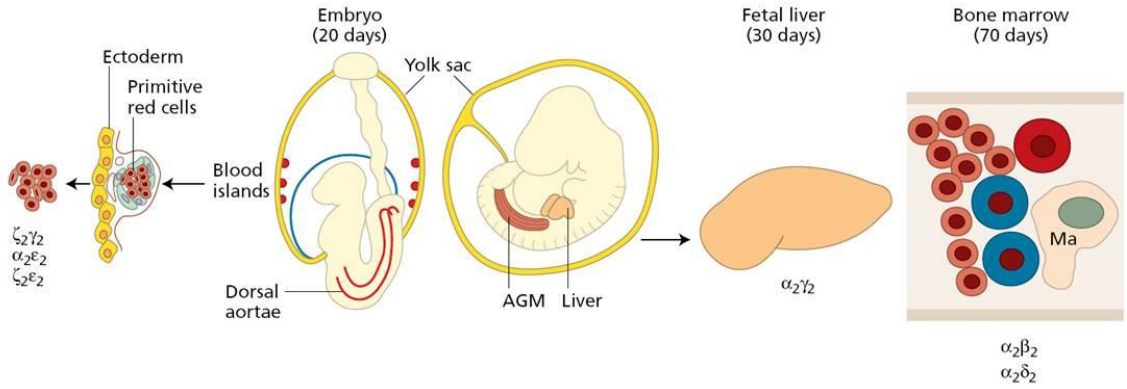


Figure 1- 6 Schematic overview of embryonic erythropoiesis.

The figure shows the formation of embryonic blood islands in the extraembryonic yolk sac and the formation of definitive haematopoiesis initially in the aorta–gonad–mesonephros (AGM) region, with subsequent migration to the liver and bone marrow. Ma denotes a macrophage. The specific types of haemoglobin formed at each stage of erythropoiesis are indicated. Adapted from Hoffbrand, 2005, p.14[2].

1.1.6 Locus Control Region (LCR)

Enhancers were discovered 30 years ago and are defined as *cis*-acting DNA regulatory elements that enhance gene transcription in eukaryotes. However, enhancers have also been found to interact with transcriptional repressors. Enhancers interact with proteins to form nucleoprotein complexes, which are able to affect chromatin structure and

activate the basal transcriptional machinery. Enhancers mediate DNA folding of clusters, which allows the physical interaction between DNA sequences separated by several nucleotides, to facilitate the interaction with general transcription factors and RNA polymerase II[20; 21].

β-globin gene locus expression depends on the *cis*-regulatory locus control region (LCR), which consists of a powerful set of enhancer elements[8] and is located from approximately 5 to 25 kb upstream (5') of the ε-gene[13; 22]. The role of the LCR in the developmental regulation of individual genes within the locus is unclear[23]. The α-globin gene cluster also contains an LCR-like region termed HS40[2]. The β-globin locus is an ideal model to study the interactions between chromatin structure and transcriptional regulation[13].

LCRs encompass a group of long-range transcriptional control determinants that regulate the expression of a large number of genes exerting a dominant transcriptional activation function over a chromatin domain[24; 25]. LCRs consists of a cluster of deoxyribonuclease (DNase) I hypersensitive (HS) sites[15]. The β-LCR, which is located 6 to 22 kb upstream (5') of the β-globin structural genes[19]. The human β-LCR is composed of 5 critical developmentally stable DNase I HSs (5'HS1 to HS5) and two weak HSs (HS6–7) [13], which represent sites of easy accessibility to interact with transcription factors and downstream gene sequences[8; 25]. HSs are 200-400 bp in size and are separated from each other by about 2-4 kbp of DNA[9]. Downstream (3') of the β-globin structural genes, another HS site (3'HS1) is present (Figure 1-5). The β-LCR is responsible of the establishment of the erythroid lineage-specific expression on the β-globin genes[10]. High-level expression within the β-globin locus requires the presence of the β-LCR[8].

The HSs of the β-LCR are specific for the cells of the erythroid lineage[26]. The β-globin locus is surrounded at both 5' and 3' ends by olfactory receptor (OR) sequences. The evolutionary conservation of the flanking elements 5'HS5 and 3'HS1 of the β-globin locus suggest the involvement of these HSs in the control of β-globin gene expression by the LCR and, at the same time, in the inactivation of the expression of the olfactory receptor genes in erythroid tissues or vice versa[12]. Each of the LCR HSs has its specific activities, with the exception of HS1 that appears to be dispensable[14]. For example, 5'

HS2 carry the enhancer activity of the β-LCR, HS3 has a domain-opening activity and functions as an activator of ε- and γ-globin gene expression. HS4 and HS5 function as insulators, being the last one contained within a matrix attachment region[10; 14; 15].

It has been proposed a model for LCR HS core formation. According to McMorrow *et al*, the first step is the binding of non-erythroid factors such as SP1 or erythroid-specific factors such as EKLF to the core forming elements, resulting in the formation of a phased nucleosomal array and a weak DNase 1 HS. Then the binding of the hematopoietic factors NF-E2 and GATA-1 occurs; the factors act to stabilize the nucleosome-free chromatin structure, indicating the formation of a strong HS. The final step involves the recruitment of non-DNA-binding factors, which interact with the DNA-bound factors, resulting in a functional HS core element[13; 27].

1.1.7 LCR and genetic activation of the β-globin locus

Chromatin opening and gene activation are separable. Activation of the β-globin locus could occur in open chromatin scenario. Thus, for the activation of the genes of the β-globin locus, an open chromatin context is necessary bu not sufficient. Chromatin opening is driven by multiple redundant elements including the LCR. Acetylation, deacetylation and other histone modifications have key roles in chromatin structures and gene expression[15]. Schübeler *et al.* reported that an additional localized H3 hyperacetylation at the LCR correlates with gene activation at the β-globin locus[28].

The preconception that the LCR is required to open or activate chromatin has been challenged by recent studies. Bender *et al.* generated mice with a targeted deletion of 5′HS1–6 in order to elucidate the role the β-globin locus of the LCR in chromatin opening and transcription during development. They reported that the deletion does not affect the open chromatin structure of the locus, remodelling of the β-globin promoter, formation of the remaining structural HS or maintenance of the nuclease-sensitive state[19]. The work of Reik *et al.* also supports this idea. They reported that deletion of the β-LCR from the native locus results in persistence of a DNase I–sensitive chromatin structure[29]. These experiments showed that the LCR is not required for activation or maintenance of an active chromatin structure in the β-globin locus, or for protecting it from a surrounding repressed chromatin. The correlate of this is that elements capable

of opening the β-globin locus by the generation of a permissive chromatin structure must exist outside of the LCR.

1.1.8 Models of LCR mediated stimulation of globin gene expression during the haemoglobin switching

Haemoglobin switching provides a model for the regulation of human gene expression during development with direct applications in the treatment of haemoglobinopathies[26]. Haemoglobin switching does not correspond to a replacement in stem cell populations expressing a fixed transcription profile of β-globin genes, but a modification in the transcription programmes of erythroid cells. Thus, the switches in the transcription of β-globin genes are controlled exclusively at the transcriptional level[10].

The human haemoglobin switch has been widely studied using transgenic mice containing the human β-globin locus (β-locus mice). These mice models are generally considered as a valid system to study human haemoglobin switching. However, recent evidence suggests that there are significant differences between humans and these mice. For instance, activation of the γ-globin genes in the transgenic mouse occurs during the embryonic erythropoiesis at the yolk sac, whereas in humans the activation of these genes occurs in the foetal liver. Another significant difference is observed during the switching from the foetal to adult globin genes, which in mice occurs during early foetal liver erythropoiesis instead of appearing around the time of birth as it occurs in humans[30]. Using these transgenic mice, various models have been proposed to explain how the LCR functions and activates the human β-globin genes in a developmental-stage specific manner: the autonomous silencing and competition models, a looping model, a linking model, a tracking model, and a sequential model. Unless stated otherwise, the switching models explained below have been developed using β-locus mice.

A) The autonomous silencing and competition models in haemoglobin switching

Expression of one gene in a locus can be negatively influenced by the transcription of other genes of the locus during haemoglobin switching. This phenomenon is known as transcriptional interference and can be explained assuming that all potentially active

genes compete for the LCR activity. Studies of human transgenes in mice showed that β-like globin genes compete for LCR interaction in a transcriptional interference context[23]. Expression suppression of the β-globin genes at the different developmental stages is thought to be achieved by at least two mechanisms. Embryonic ε- and foetal γ-globin genes autonomously suppress their expression. However, the β-globin gene silences its expression during the embryonic stage in a competitive fashion for the LCR with the other genes of the locus[17].

Two models have been proposed to understand how the LCR is involved in the expression of β-globin genes. On the one hand, activation of the β-globin genes is conducted by an environment of open chromatin provided by the LCR. Thus, β-globin genes would be autonomously regulated. On the other hand, a competitive model is proposed to explain the sequential activation of the particular β-globin genes. According to this model each gene of the β-globin locus competes for the LCR, but only one gene-LCR interaction would be possible per chromosome[17]. The results of such competition might depend on the array of transcription factors interacting with both gene promoters and the LCR. Whereas the autonomous model seems to be employed more frequently, literature supports the competition model to explain the foetal silencing of the β-globin gene, which competes with the y-globin gene for the LCR activity[7].

The gene competition is the prevalent model for the developmental silencing of the adult β-globin gene[23]. This model is supported by the work of Behringer *et al.*[18] and Enver *et al.*[31]. This group worked with transgenic mice to understand the mechanism of γ- to β-globin gene switching. They observed that during the foetal stage there is a preferential interaction between the γ-globin genes and the LCR, and the β-globin gene is turned off competitively. In contrast, in the adult stage the LCR interacts preferentially with the β-globin gene whereas the γ-globin genes are silenced. During embryonic erythropoiesis, the β-globin gene is silenced by the competition by the upstream ε- and γ-globin genes for the LCR (Figure 1-7)[10].

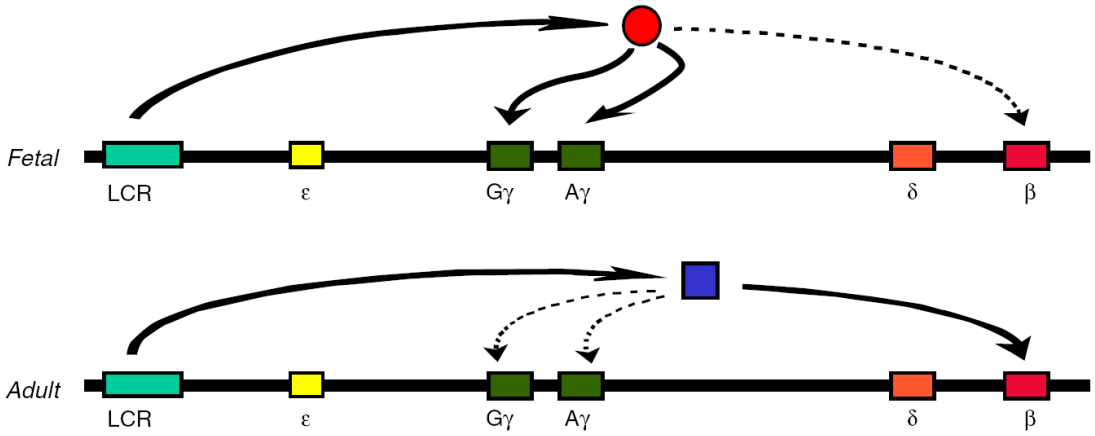


Figure 1- 7 The competitive mechanism of haemoglobin switching.

There is a competition between the adult and foetal globin genes to interact with the LCR. During foetal life γ-globin genes interact with the LCR (red circle) and β-globin gene is slenced (diffuse arrow). In the adult stage, the β-globin interacts with the LCR (blue rectangle) whereas the γ-globin genes are silenced (diffuse arrows). Adapted from: Stamatoyannopoulos, 2005, p.264[10].

The autonomous silencing model has been adopted to explain the silencing of the embryonic ε-globin gene, which expression is totally restricted in yolk sac erythropoiesis and completely dependent on the presence of the LCR[32].

B) Models for LCR-β-globin gene interaction

(i) Looping model

Choi *et al.* first proposed the looping model assuming that a loop through the nucleoplasm is formed by the direct interaction of distal and proximal regulatory elements to enhance the expression of the individual genes of the β-globn locus[33]. The looping model suggests that activation of β-globin genes in erythroid cells is preceded by a loop between 5'HS5 and 3'HS1, delineating a β-globin chromatin domain[12] (Figure 1-8). It has been proposed that each HS of the LCR, and the transcription factors bound to them, contribute cooperativitely to form a larger physical synergistic entity termed holocomplex[13; 15]. This complex interacts directly with the genes to activate its transcription[17]. The looping of this complex allows the association of the LCR with gene-proximal elements that activate the transcriptional machinery[10]. Carter *et al.*, using

a so-called RNA TRAP (tagging and recovery of associated proteins) showed that the 5'HS2 of the β-LCR is very close to the β-globin gene located when this is active, suggesting a direct regulatory interaction[34] (Figure 1-8).

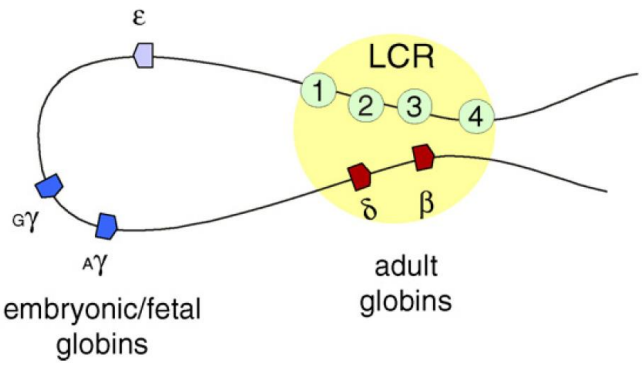


Figure 1- 8 Looping model.

The β-globin locus LCR and adult β-globin gene come in to close proximity in adult stage erythroid cells in which that gene is actively transcribed. The embryonic and foetal expressed genes that are silent at this stage of development are excluded from the close contacts. Adapted from: Kiefer, 2008, p.69[12].

(ii) Linking model

According to the linking model, the LCR activates individual gene promoters by leading the formation of a chain of components of the transcriptional machinery that extends in the promoter of the gene to be activated[15].

(iii) Tracking model

Tuan *et al*. proposed the tracking model from the observation that the HS2 of the LCR starts the production of long non-coding RNAs. According to their model, the LCR recruits transcription complexes to the β-globin gene promoters[35].

(iv) The facilitated-tracking model

In the facilitated-tracking model are combined different aspects of the looping and tracking models. According to this model, transcription factors bind to the HSs of the

LCR forming a holocomplex that loops to track downstream DNA, and transcription factors are released in specific promoter regions[15].

(v) Sequential model

Schübeler *et al.*[28] proposed a sequential model for the β-globin gene activation based on the dissociation of chromatin opening and transcriptional activation. The first step of the model is the relocation, by *cis*-elements others than the LCR, of the β-globin locus in a region free of centromeric heterochromatin to achieve the propagation of an open chromatin context. The establishment of an open chromatin structure is marked by broad acetylation, which involves the formation of HSs at the LCR and the β-globin gene promoters. The next step consists in the localized H4 hyperacetylation at the active promoters, process in which the LCR plays an active role. LCR-mediated high-level transcription is the final step of the model. The LCR modifies or recruits additional components to the remodelled and hyperacetylated promoter. Alternatively, LCR may help the elongation efficiency of transcription initiated at the β-globin gene promoter[13].

1.1.9 Haemoglobinopathies associated with the β-globin locus

Blood disorders are amongst the most common genetic diseases in the world and have been associated with serious, long-term disability, lower-than-average life expectancy and severe morbidity[6]. These inherited disorders have been used as a models to study the developmental regulation of the β-globin locus and haemoglobin switching[17]. Haemoglobinopathies are characterized for the absence of functional α-like or β-like globin chains[6]. In addition, a harmless group of mutations known as hereditary persistence of foetal haemoglobin also have been shown to interfere with the normal haemoglobin production[1].

A) Thalassaemias

Thalassaemias correspond to the pathological state characterised by the defective production in some of the chains that conform the haemoglobin molecule. Thalassemias are divided in function to which globin chain is produced in reduced amounts, that is α-, β-, δβ- or γδβ-thalassaemias. In α^0- or β^0-thalassemias no globin chain is synthesized at all[2].

(i) β-Thalassaemias

The most common and relevant form of thalassemias are those in which no β-chain is produced (β-thalassaemias), leading to a severely decreased or absent adult β-globin production, which is associated severe anaemia[2]. β-thalassaemias are generally caused by point mutations, minor insertions or deletions, which affect different levels in the β-globin synthesis cascade[36]. In β-thalassaemia, the synthesis of α-globin continues as normal, resulting in its accumulation to finally precipitate in the erythroid precursors. This causes the formation of inclusion bodies which lead to the premature destruction of the erythroid precursors in the bone marrow, a phenomenon known as ineffective erythropoiesis[8].

B) Sickle cell disease

Sickle cell disease is the consequence of the replacemant of a thymine for an adenine in the sixth codon of the β-globin gene. This change in the genetic sequence leads to the replacement of valine for glutamic acid at the sixth position of the chain of the β-globin. When HbS deoxygenates the conformation of the molecule changes, exposing the hydrophobic β^6 valine at the surface of the β^s-globin chain. This acts as a binding site for another complentary haemoglobin tetramer, triggering the formation of large polymers. These tetramers rapidly form relatively large fibers, which deform red cells upon deoxygenation due to the polymerization of the defective globin tetramers. HbS polymers reduce red cell deformability, which leads to increased blood viscosity. Sickling also damages the red cell membrane, which finally leads to haemolytic anaemia, and clogging of small vessels(Figure 1-9)[2; 8].

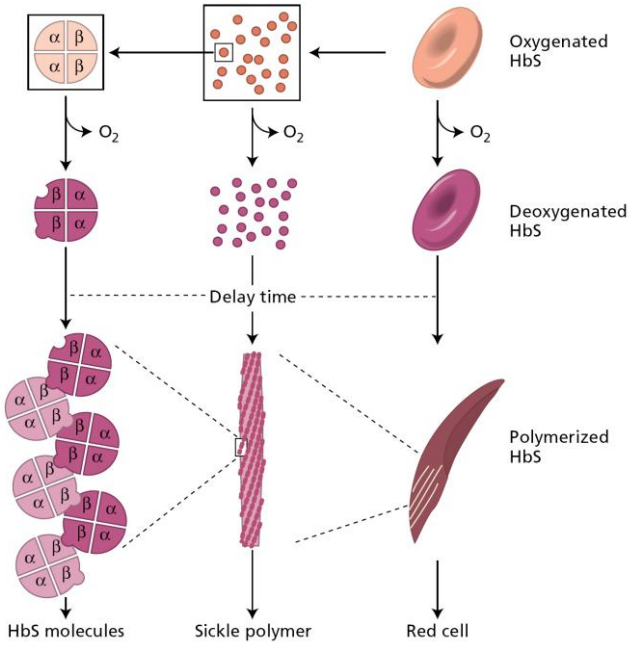


Figure 1- 9 Induction of red cell sickling.

As red cells traverse the microcirculation, oxygen is released from oxy-HbS (red circles), generating deoxy-HbS (purple circles). Conformational change exposes a hydrophobic patch at the site of the β^6-valine replacement, shown as a projection (left column), which can bind to a complementary hydrophobic site on a subunit of another haemoglobin tetramer, shown as an indentation. The middle column shows the assembly of deoxy-HbS into a helical fibre, shown as a twisted rope-like structure. As deoxy-HbS polymerizes and fibres align, the red cell is distorted into sickle shape (right column). Source: Hoffbrand, 2005, p.106[2].

C) Hereditary persistence of foetal haemoglobin

Hereditary persistence of foetal haemoglobin (HPFH) is a heterogeneous group of asymptomatic conditions characterised by persistent foetal haemoglobin production during adult life. HPFH is caused by long deletions of the β-globin locus and localized mutations in the upstream promoters of the $^{G}\gamma$ and $^{A}\gamma$-globin genes. These changes allow the γ-globin genes to be active in adult life. Despite of his nule clinical symptoms, its importance lies on his potential to modify the phenotype of the β-haemoglobinopathies and sickle cell disease, since it causes a high level of HbF

production and hence reduces the severity of the phenotype. Thus, HPFH has been proposed as a potential therapy to fight these diseases[2].

D) Therapeutic approaches to reactivation of γ-globin and HbF in adult erythroid cells

A full comprehension of the molecular mechanisms of globin switching is appears to be necessary to develop new therapies for β-thalassaemia or sickle cell disease based on the reactivation of γ-globin and HbF in adult erythroid cells. Indeed, treatments for red blood disorders aim, somewhat unsuccessfully, to reactivate the foetal version of haemoglobin. HbF production can be increased by erythropoietin, other cytokines, and changes in cell cycle kinetics. Butyrate or hidroxyurea-based treatments can increase HbF levels in patients suffering from β-thalassemia and sickle cell disease[8], fuelling interest in the reactivation of γ-globin as a therapeutic measure[10]. Characterization of γ-globin gene regulation and human haemoglobin switching will be useful to develop new therapies to reactivate γ-globin expression in adult erythroid cells[12].

1.2 ncRNAs & Chromatin

1.2.1 Non-coding RNAs

Most of the genome of mammals, and most of the eukaryotes, is transcribed to produce coding and non-coding RNAs (ncRNAs)[37]. ncRNAs are a growing class of RNAs with biological functions rather than simple intermediate messengers between DNA and proteins. ncRNAs usually do not encode for Open Reading Frame (ORF), but contain a high number of stop codons[38]. Several pieces of evidence support that ncRNAs are versatile molecules that, due to their chemical properties, can form complex tertiary structures that allow them to perform roles that were thought to be exclusive of proteins[39]. The abundance of ncRNAs is a distinguishing feature of metazoan genomes[40]. ncRNAs are a heterogeneous group which is divided into three subgroups according to their length and function[38]: ncRNAs range from about 1) ~18-25 nucleotides for microRNAs, ~20-300 nucleotides for small RNAs commonly found as translational regulators, and up to and beyond 10,000 nucleotides for large RNAs involved in gene silencing and other processes[39] (Figure 1-10).

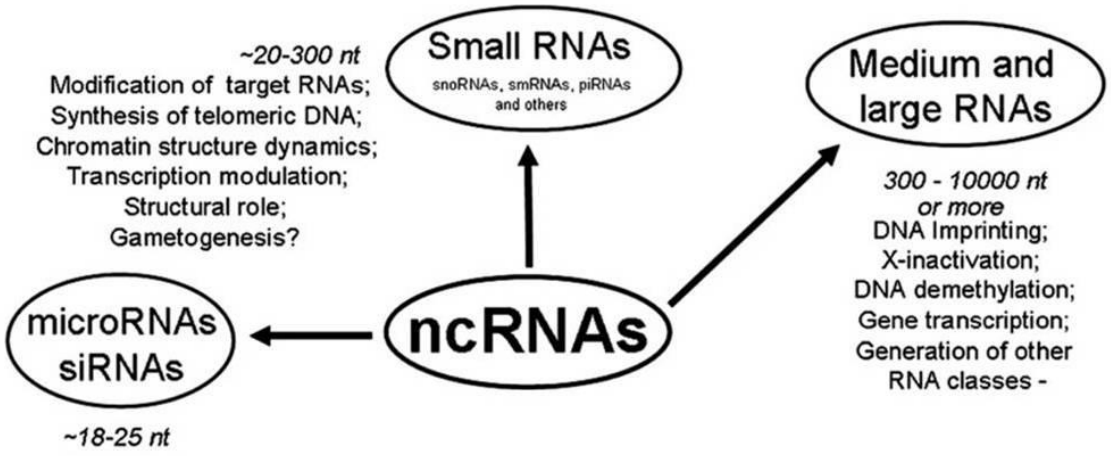


Figure 1- 10 Schematic representation of the emerging ncRNA world.
microRNAs, siRNAs, small RNAs, and medium/large RNAs. Adapted from: Costa, 2007, p.2[38].

In the past few years, several studies[39] have demonstrated that the biological functions of ncRNAs have been largely underestimated by the emphasis on mutation scanning of protein-coding exons[37]. Evidence from several studies using a variety of model systems demonstrates the key role of ncRNAs during cellular differentiation and organism development, through a variety of complex mechanisms such as the recruitment of proteins, the control of epigenetic regulation, chromosomal and chromatin structure dynamics, long-range interactions, gene silencing, gene transcription, DNA imprinting, DNA methylation and RNA interference[38; 41; 42].

ncRNAs can control gene expression at the levels of translation, transcription, and post-transcriptional processing, affecting mRNA stability or processing, including the regulation of alternative splicing[43]. ncRNAs not only carry out different functions in the nucleus, but also function in dosage compensation and imprinting, and in other type of gene regulation, either in *cis* or in *trans*[40; 44].These new evidences challenge the assumption that protein-coding genes are also the main controllers of cellular processes, suggesting a wide-spread involvement of ncRNAs in regulation[45]. All these data together suggest that the transcribed portions of genomes are larger than what was thought, where ncRNAs represent more than 50% of cellular RNA[41], and that a huge part of the epigenetic regulation relies in these untranslated transcripts[46].

The complexity of the transcriptome has prompted a shift from the simplistic understanding of a linear genomic organization to a modular model, which includes coding and non-coding transcripts[37]. ncRNAs have been shown to be part of complex regulatory networks, suggesting that the simplistic protein-centric view of the mammalian genome is wrong and that intergenic transcription is functional and not just transcriptional noise[43]. Data from different studies suggest that almost 90% of the non-repetitive sequences of the mammalian genome are transcribed, being a large proportion of these ncRNAs, originating the alleged paradox that ncRNAs represent the largest output of the human genome. These intergenic transcripts also play a crucial role in epigenetic transmission of early determined transcription states[44]. It has been shown that complex organisms have a greater amount of regulatory information in their genome. This is supported by the observation that in bacteria regulatory factors scale quadratically with genome size[47]. These regulatory factors include the untranslated regions (UTRs) and non-protein-coding intronic and intergenic sequences, which increase dramatically in size with increasing developmental complexity in eukaryotes. This observation adds to the list of evidences about the regulatory function of ncRNAs in gene expression during differentiation and development[43].

Most ncRNAs show cell- or tissue- and temporal-specific expression. Furthermore, altered expression patterns of ncRNAs are involved in tumour-associated phenotypes and other diseases[43]. ncRNAs have been shown to take part in differentiation and specific developmental processes in multicellular organisms, acting as both activators and repressors of gene expression in complex regulatory circuits integrated at various levels[43]. ncRNAs show lower expression levels and lower sequence conservation than mRNAs, raising the possibility that ncRNAs arise by illegitimate transcription and are no functional[37; 43].

1.2.2 ncRNAs and Chromatin Regulation

Chromatin is composed of nucleosome subunits, which are nucleoprotein complexes formed by histone octamers (two copies of each of the four core histones H2A, H2B, H3, and H4) that enclose a specific DNA sequence of 147 base pairs in length[42; 48]. Chromatin is highly compacted, allowing to pack 2 metres of DNA in an eukaryotic nucleus. The problem is that this compacted structure prevents the access of

transcription factors to DNA. Thus, during gene expression, a regulated disruption and reconstitution of the chromatin structure is needed[48].

Chromatin structure can be dynamically altered by methylation of the DNA[43], and by enzymes that disrupt histone-DNA contacts in an ATP dependent way to allow the access of the transcriptional machinery to the nucleosomal DNA[48]. Posttranslational modifications of the core histone proteins, which can be acetylated, methylated or phosphorylated, epigenetically regulate gene expression. For example, Polycomb-group proteins catalyze the methylation of H3 lysine 27 (lys27), a modification that promotes the compacting of chromatin that in turn negatively regulates transcription[49].

It has recently become apparent that transcription affects the nature of chromatin, a concept that contrasts with the established idea that transcription is affected by chromatin structure[16]. ncRNAs have been shown to be associated with eukaryotic chromatin and play an important role in its structure maintenance[39]. Non-coding transcription correlates with chromatin structural alterations and locus activation of several loc in mammals, which often show developmental regulation. This suggests the involvement of intergenic transcription in the establishment of open chromatin domains[50; 51]. For example, intergenic transcription correlates with histone modification and with domains of highly modified chromatin[50].

Non-coding transcription by RNA polymerase II (RNAPII) occurs across intergenic regions, introns, and exons. Many cases of intergenic transcription associated with promoters have been reported, suggesting that intergenic transcription across promoter has a regulating function during gene transcription[52]. As discussed below, different types of ncRNAs act as recruiters to specific loci of DNA methyltransferases and chromatin-modifying complexes, evidencing the involvement of ncRNAs in the modulation of chromatin structure dynamics[43].

1.2.3 Examples of ncRNAs in the regulation of gene expression

A) Air ncRNA

Genomic imprinting is an epigenetic phenomenon[53] characterised by monoallelic expression of genes in a parent-of-origin dependent manner[54-56]. The imprinting control

region (ICR) *H19/Igf2r* DMR (insulin-like growth factor 2 differentially methylated region) gene is one of the most extensively studied ICRs[54]. The key regulatory elements for the imprinted expression on *H19* and *Igf2r* from the paternal and maternal alleles respectively, are the ICR upstream of *H19* and a set of enhancers downstream of the *H19*[57]. Thorvaldsen *et al.* found that mouse with a deletion in this sequence lost the imprinted expression of the *Igf2r* and *H19* genes[57]. Schoenfelder *et al.* reported that the *H19* ICR is biallelicaly transcribed in both sense and antisense directions. They were the first group in report the functionality of ncRNAs in the *H19* ICR, indicating their key roles in the regulation of gene expression[53].

The long *Air* (Anti-sense IGF2r RNA) transcript is initiated from the *Igf2r* locus in the intronic differentially methylated region (DMR) Region 2[58], regulated by DNA methylation specific to the parent of origin[59]. Wutz *et al.* reported that deletion of Region 2 results in the biallelic expression of *Igf2r*, *Slc22a2*, and *Slc22a3*, three imprinted genes exclusively expressed only from the maternal alleles[60; 61]. *Igf2r* is overlapped by *Air*, which is transcribed in the antisense direction through the *Ifg2r* locus and its expression is imprinted from the paternal allele exclusively. Imprinted repression of *Igf2r*, *Slc22a2* and *Slc22a3* requires transcription of *Air* beyond the second exon of *Igf2r*[62]. Two models were proposed by Sleutels *et al.* to explain the mechanism of gene silencing by *Air*: 1) Antisense transcription across the *Igf2r* locus to repress the promoter of the overlapping gene, inducing a silent chromatin state that could spread bidirectionally in a limited manner into *Slc22a2* and *Slc22a3*, and 2) *Air* forms a ribonucleoprotein to coat and silence flanking chromatin of the *Igf2r* locus[63].

Nagano *et al.* reported an interaction in placenta between *Air* the *Slc22a3* promoter and the H3K9 histone methyltransferase G9a. They reported that there exists a correlation between the accumulation of *Air* at the *Slc22a3* pomoter and localized H3K9 methylation and transcriptional repression of the *Slc22a3* gene.

They concluded that *Air* epigenetically silences the *Slc22a3* gene via the recruitment to the *Slc22a3* gene promoter of G9a, which leads to targeted H3K9 methylation and allelic silencing[59]. All these data together evidence that the *Air* ncRNA is directly involved in lineage-specific transcriptional repression[64], based on the recruitment of repressive histone-modifying activities to epigenetically silence transcription[59].

B) HOTAIR ncRNA

The *HOX* gene clusters are a classic model for studying gene regulation during embryonic development[65]. Thirty-nine *HOX* transcription factors have been reported in mammals. They are clustered in four chromosomal loci consisting of 9 to 11 *HOX* genes arranged in tandem[66], termed *HOXA* through *HOXD*. It has been described a complex epigenetic regulation behind the maintenance of *HOX* expression patterns. Their complex clustering on the chromosomes and the large number of *HOX* ncRNAs, suggest that ncRNAs have a key role in *HOX* genes transcriptional *cis* regulation[40]. *HOX* genes are present in all bilaterian animals and their genomic organization is highly conserved[66]. These transcription factors have central roles in embryonic patterning and in developing of adult organs, and have been shown to be essential in the specification of the positional identities of cells[66]. Hundreds of ncRNAs are transcribed from the regulatory regions of *HOX* clusters and have been shown to be involved in Polycomb-mediated regulation, counteracting Polycomb repression[43]. The Polycomb group proteins (PcG) work as large complexes that act on nucleosome structure and the RNA–Polymerase II complex, leading to the inhibition of transcriptional activation. PcG complexes are involved in the control of developmental regulators, including all four *HOX* clusters[44]. These complexes are associated with trimethylated K27 on histone H3, a modification frequently required for this epigenetic silencing[65]. The initiation of histone modification is carried out by the Polycomb Repressive Complex 2 (PRC2) which consists of a H3K27 histone methyl transferase (HMTase) EZH2 and the core components Suz12 and EED, whereas the maintenance of these modifications and the promotion of chromatin compaction is done by the Polycomb Repressive Complex 1 (PRC1)[40].

Sessa *et al.* showed that opposite strand-intergenic transcripts are coexpressed with adjacent coding *HOXA* genes. They concluded that the *HOXA* cluster shows a highly complex profile of non-coding transcripts, and that non-coding transcription is essential for the opening and maintenance of the active state of *HOX* clusters. They also suggested that PcG proteins have a role in *HOX* gene regulation[44]. Their results are in consonance with the work of Bernstein *et al.*, which also concluded that the

transcription of intergenic sequences, which represent methylated and transcriptional active chromatin domains, are needed for the maintenance of *HOX* gene expression[67].

Rinn *et al.* identified 231 long ncRNAs in the four *HOX* human clusters by combining the powers of ultrahigh-resolution tiling microarrays and bioinformatics algorithms in a collection of primary human fibroblasts with 11 distinct positional identities[40]. They also showed that these ncRNAs are collinearly expressed along developmental axes and share some common sequence motifs[43]. Most of these transcripts are intergenic, are transcribed in the direction opposite the *HOX* genes, and are differentially expressed in different cell types in function of their original position along the body axis[65]. One of the ncRNAs identified by Rinn *et al.* showed a surprising feature. *HOTAIR* acts in *trans* interacting with PcG proteins and recruits the Polycomb complex to the *HOXD* region where it trimethylates lysine 27 residues (me3K27) of histone H3 to induce heterochromatin formation and repress gene expression[37](Figure 1-11).

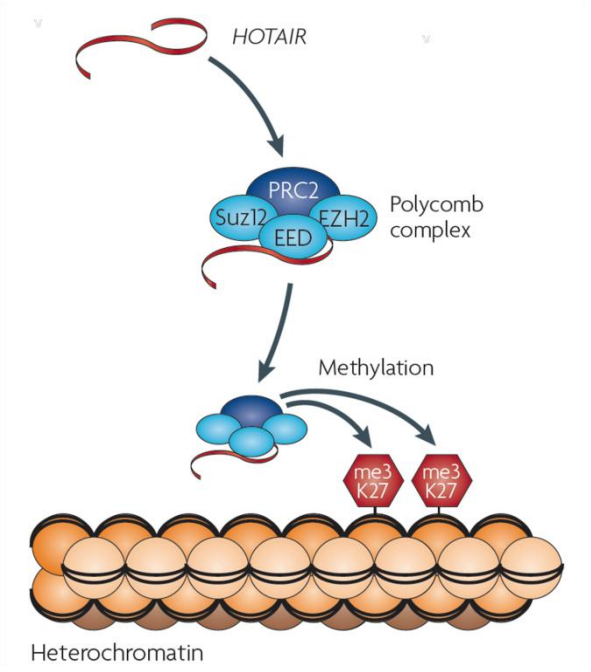


Figure 1- 11 Chromatin remodelling by *HOTAIR*

HOTAIR recruits the Polycomb complex to the *HOXD* locus where it trimethylates lysine 27 residues (me3K27) of histone H3 to induce heterochromatin formation and repress gene expression. Adapted from: Mercer, 2009, p.158[37].

This observation of a *HOX* ncRNA binding to an epigenetic regulatory complex to change methylation patterns on a different *HOX* cluster on a different chromosome was unexpected[65]. The characterization of *HOTAIR* raises the question if are there other *trans*-acting ncRNAs from the *HOX* cluster or elsewhere that have not yet been found[65].

C) ncRNAs in X chromosome dosage compensation

The X chromosome carries genes required for both sexes, whereas the Y chromosome has little information. As many animals have a single X chromosome in males (XY karyotype) and two in females (XX karyotype), it results in an imbalance of X-linked genes in one sex[68]. The balance between X-chromosome and autosome expression in both sexes is achieved through an epigenetic regulatory mechanism termed dosage compensation, which modulates global expression of an X chromosome. This modulation, which is very important for proper development[39], is achieved by directed chromatin modifications targeting one of the X chromosomes[68]. Normalization of the copy number of X chromosome between male and female cells[42] by dosage compensation is one of the best examples for understanding the role of ncRNAs in the formation of euchromatin versus heterochromatin. ncRNAs are intimately involved in dosage compensation in both *Drosophila* and mammals. Long ncRNAs, such as *Xist* in mammals and *roX* in flies, have been shown to have key roles in X-chromosome dosage compensation[68], which spread in *cis* to coat the X chromosome[42]. In both systems X chromosome dosage compensation by ncRNAs leads to changes in chromatin structure and act as recruiting entities or scaffolding factors in the cooperative binding of chromatin-associated complexes[42], but different strategies are used[43]. In flies this epigenetic process is characterized by an up-regulation of almost twofold of most of genes on the single X chromosome in males, whereas in mammals consists in the silencing of most genes in one of the two X chromosomes in females[42].

(i) X chromosome dosage compensation in mammals

X chromosome inactivation (XCI) in mammals involves silencing of one of the two X chromosomes in females. This process is controlled by the X-inactivation centre *(Xic)*, which counts the number of X chromosomes, chooses one to remain active and silences the inactive X chromosome. *Xic* contains three ncRNAs essentials for this process: X

inactive-specific transcript (*Xist*) and *Tsix*, which form a sense-antisense pair, and *Xite*. One of the early steps of XCI consists in the sequential recruitment of two Polycomb complexes to repress transcription by establishing repressive histone methylation marks. The point of no return in XCI, i.e. when the X chromosome silencing is irreversible is achieved some days later in a process that involves DNA methylation[43]. XCI randomly selects the X chromosome to be silenced in females when *Xist* is exclusively transcribed from the future inactive X chromosome. *Xist* silences in *cis* the selected X chromosome by coating him and recruiting Polycomb-group proteins to establish and spread heterochromatin[43]. This process is achieved via trimethylation of histone H3 on lysine 27 (H3K27me3), rendering the chromosome transcriptionally silent in a similar way to *HOTAIR*[40]. It has been observed a correlation between the coating by *Xist* of the entire chromosome that will become the inactive X (Xi), and the establishment and maintenance of the transcriptional silencing along the Xi[42; 62]. *Tsix* ncRNA, which is transcribed in the antisense direction through the *Xist* locus, negatively regulates *Xist*. The key role of *Tsix* consists in decide which X chromosome will become the active X (Xa) and the Xi by overlapping the full length of *Xist*[62].

(ii) X chromosome dosage compensation in fruit flies

In *Drosophila melanogaster* X chromosome dosage is addressed by doubling the transcription rate from the single X chromosome in XY males relative to XX females[62] by a ribonucleoprotein complex that binds to hundreds of sites along the male X chromosome. This complex, termed Male-specific lethal (MSL) complex, consists of five core proteins: MSL1, an acidic protein that can interact with MSL2, MSL3 and MOF through distinct domains; MSL2, a Ring-finger protein that regulates assembly of the complex and is only expressed in males; MSL3, a chromodomain protein that carries a MRG [Mortality factor on chromosome 4 (MORF4)-related gene] domain necessary for X localization; Maleless (MLE), a RNA helicase; Males absent on first (MOF), a chromodomain-containing histone acetyltransferase. In addition a H3 Ser 10 kinase known as JIL-1 interacts with the MSL proteins[42; 68] (Figure 1-12).

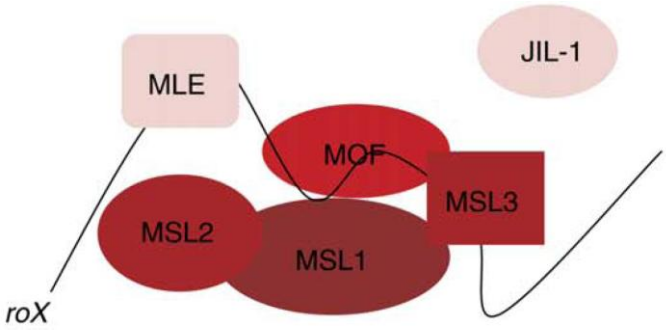


Figure 1- 12 The MSL complex.

The MSL complex comprises five proteins and *roX* RNA and binds to hundreds of sites along the male X chromosome to mediate dosage compensation. MSL1, MSL2, MSL3 and MOF interact directly with each other. MLE and JIL-1 are enriched on the male X chromosome but are thought to have a more peripheral association and also bind to autosomal sites. Adapted from: Deng, 2006, p.527[68].

The function of MSL is to spread hyperactive chromatin along the X chromosome. This process is ATPase dependent for the activity of the MEL helicase and the acetyl transferase activity of MOF[62]. MOF is the key regulatory component of the MSL complex, which hyperacetylates histone H4 at Lys16, resulting in the up-regulation of transcription from the X chromosome[62]. It has been suggested that Lys 16 acetylation either could help to recruit a positively acting effector to the male X chromosome or to displace a silencing factor[42]. The *roX1* and *roX2* ncRNAs, encoded on the X chromosome, are also components of the MSL complex[62]. These ncRNAs are incorporated into the MSL complex by MLE. Both *roX* ncRNAs act in *cis* on the male X chromosome from which are transcribed[62] by guiding the localization of the MSL proteins to hundreds of sites on the male X chromosome[68]. This process is thought to be guided by base-pairing between the *roX* RNAs and DNA. Thus, *roX* ncRNAs enhance the targeting and the processivity of MSL[62]. Unlike *Xist* ncRNA in mammals, *roX* ncRNAs direct activation, rather than silencing, of target genes. These RNAs have key roles in the epigenetic establishment of correct genome expression by a mechanism of RNA-directed chromatin regulation[68].

D) SRG1 ncRNA

Martens *et al.* reported that in *Saccharomyces cerevisiae* the repression of the *SER3* gene, which encodes a phosphoglycerate dehydrogenase that catalyses a step in serine biosynthesis, is mediated by regulated transcription across *SER3* from the upstream *SRG1* promoter. They observed that the regulatory region of *SER3* gene was highly transcribed and produces a ncRNA called *SRG1* when the yeast is grown in rich medium. The repression of *SER3* transcription by *SRG1* is driven by a transcription-interference mechanism in which the transcription of the ncRNA across the *SER3* promoter interferes with the binding of activators to the promoter[69].

E) ncRNAs and disease

Evidence that ncRNAs are implicated in diseases come from the work of Sonkoly *et al.* and Lu *et al*[39]. The first group identified a long ncRNA involved in psoriasis termed *PRINS* (Psoriasis susceptibility-related RNA Gene Induced by Stress)[70]. Lu *et al.* showed that the long ncRNA *KLK3IP* was associated with the prostate cancer[71].

F) Remarks of the importance of ncRNAs

Taken together, the data of the studies described here are just the tip of the iceberg of the broadly used mechanism of intergenic transcription for transcriptional regulation[72]. In the near future bioinformatic tools will generate more accurate databases of predicted ncRNAs to be used to describe new ncRNA examples[39]. Deletion, over-expression or down-regulation have been shown to underlie and act as markers for the severity of several complex diseases[39]. It is becoming clear that ncRNAs are essential molecules in the complex network of interactions that controls gene expression and they are not simply transcriptional noise or evolutionary junk[38]. The phenomenon of whole-genome transcription in metazoa and probably all eukaryotes, suggest that genome physiology is regulated in a developmental manner by ncRNAs that orchestrate the precise patterns of gene expression during morphogenesis[43]. Genetic programming of complex eukaryotes has been evolutionary linked with the expansion of ncRNAs, as it is observed in the concomitant increase in non-coding content with organism complexity[37].

1.2.4 Chromatin structure and remodelling in the control of β-globin locus expression

The onset of erythropoiesis in erythrocyte precursors and the differential expression of the different genes of the β-globin locus during development correlate with wide epigenetic modifications in the locus[10]. In the β-globin locus, as in other structural gene loci, acetylation of histones in chromatin leads to increased expression of the locus genes, by making gene sequences more available to transcription factors. The human β-globin locus shows a complex pattern of histone modifications, extremely different between foetal and adult erythrocytes. There is a correlation between H3 and H4 acetylation and H3K4 methylation, which are modifications associated with genetic activation, and DNase I hypersensitivity sites in the β-LCR[50]. During differentiation, within the β-globin locus the active genes and the LCR acquire positive histone modifications. In contrast, inactive genes are marked with silencing histone [12] (Figure 1-13).

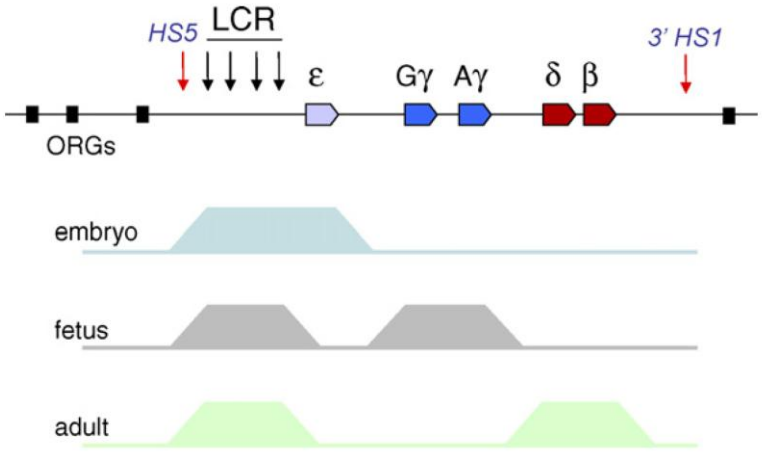


Figure 1- 13 Epigenetic changes during development of the β-globin locus.
The coloured boxes indicate domains of H3K4 di-methylation and histone acetylation during different developmental stages of the β-globin locus. Source: Kiefer, 2008, p.73[12].

When the LCR interacts with the proximal promoter of one of the β-globin genes, it recruits histone acetyl transferases (HATs) to acetylate histones. Then, the transcriptional machinery binds to the hyperacetylated promoter and the gene

transcription is activated. When the individual gene is to be inactivated, the acetylation level is decreased, leading to the spread of repressed chromatin structure. The sequential activation of the individual β-globin genes during development is attributed to this propagation[15].

1.2.5 Intergenic transcription in the β-globin locus

Although the human β-globin locus is amongst the most extensively characterized gene clusters in the human genome at the gene and protein levels[8; 12], the extent of transcription throughout the locus remains to be studied[25]. Non-coding RNAs span across the human β-LCR and intergenic regions of the β-globin locus in erythroid cells[12], but at a lower abundance than the gene transcripts[50]. Intergenic transcription of the β-globin locus may be ligated to β-LCR activity if it is assumed that the function of transcription establishes and maintains an open chromatin context within the locus. However, intergenic transcription within the locus could just be the consequence of its open chromatin context. The arguments against this last assertion are based on the apparent strand specificity of transcription and specificity of the sites at which transcription initiates in the intergenic regions[24; 25].

Chromatin structure can be affected by transcription, since RNA polymerase II (RNAP II) complexes exhibit chromatin-remodelling activity. The elongating form of RNA polymerase II is associated with histone modifying and chromatin remodelling activities that could be responsible of modified domains observed in the regions of intergenic transcription within the β-globin locus. Thus, intergenic transcription could control the increase in histone acetylation initiated from the LCR in the β-globin locus via RNA polymerase II.

The precise role of intergenic transcription in the β-globin locus remains largely unclear, but it has been suggested that the establishment of active chromatin domains and the transcriptional regulation of the locus are mechanistically related to intergenic transcription and histone modification[12]. One of the roles of the β-LCR and intergenic transcripts may be to deliver proteins and/or RNA polymerase II to the globin gene promoters, as it is proposed in the tracking model[25]. There is a good correlation between the acetylation of the different subdomains of the locus and intergenic transcription at

different developmental stages, suggesting that these non-coding transcripts are developmentally regulated[12].

A) Evidence for intergenic transcription within the β-globin locus

In 1992 Tuan *et al.* proposed the tracking model for haemoglobin switching based on the observation that in addition to genic transcription in the β-globin locus, LCR HS2 initiates the formation of long non-coding transcripts which transform the chromatin conformation of the individual genes of the locus to an open context to activate their transcription. According to their model the LCR recruits transcription complexes that track along the DNA until they reach the β-globin gene promoters[35]. The work of Tuan *et al.* was supported by the experiments of Johnson *et al.,* who demonstrated that RNA polymerase II (Pol II) binds the β-LCR far upstream of the β-globin promoters, concretely to HSs within the LCR[73].

Tuan *et al.* have recently further analyzed the HS2 enhancer-initiated RNAs and their mode of transcription from the HS2. They suggested that the generation of these non-coding intergenic RNAs could help to facilitate tracking of the transcriptional complex through the intervening DNA to reach the basal promoter complex and activate efficient mRNA synthesis from the promoter[74].

It could be concluded from this data that intergenic transcription could have a role in propagating DNAse sensivity in the β-globin locus initiated by the binding of transcription factors, and thus be correlated to chromatin opening within the locus[16]. This conclusion is supported by the work of Reik *et al.,* who generated a targeted deletion of 5'HS2 to HS5 in the human β-LCR. Although globin gene expression was disrupted by this deletion, the β-globin locus retained its erythroid-specific DNase I sensitivity[75].

Ashe *et al.* analyzed the nascent transcription across the human β-globin locus in erythroid and non-erythroid cells. They reported that is possible to induce intergenic transcription from the β-globin locus in non-erythroid cell lines by transient transfection of a plasmid containing an actively transcribed gene of the locus. In contrast, the genic transcripts of the β-globin locus were not induced due to the absence of specific erythroid transcription factors[25].

B) Distinctive histone methylation patterns in transcribed coding and non-coding human β-globin sequences

Kim *et al.* studied the differences in histone modifications between the coding and non-coding regions within the β-globin locus. They found that inactive genes and intergenic transcription were broadly marked by K4, K9, and K36 monomethylations. However, the coding regions of the highly transcribed genes were H3K4, K9, and K36 trimethylated. These observations indicate that the transcriptional activation of the genic and intergenic transcription is modulated by independent mechanisms[76].

C) Intergenic transcription in relation to development

Gribnau *et al.* investigated LCR and intergenic transcription and chromatin structure during development in the human β-globin locus. They showed that the β-globin locus is divided into three differently active subdomains and they can change their status according to the activated gene of the locus at different developmental times: the embryonic and foetal εγ subdomain, the foetal and adult δβ subdomain, and an LCR subdomain are active at all developmental stages (Figure 1-15). They reported that the production of intergenic transcripts in each case is congruent with increased DNase I sensitivity and in the case of the εγ and δβ subdomains coincided with the activities of the genes themselves[77].

This group investigated the extent of intergenic transcription in the β-globin locus during development using RNA fluorescence *in situ* hybridisation (FISH) with multiple gene-specific intron probes (a-g) targeted to transcripts of intergenic regions of the β-gobin locus (Figure 1-14). Their results showed that intergenic transcripts are developmentally regulated and delineate large regions of 15–30 kb surrounding active genes. They also observed that transcripts in the εγ and δβ subdomains are independently regulated. Probes c, d and e targeted intergenic transcripts which expression was predominant during embryonic stage, but decreases in adult stages of development. In the other hand, transcripts targeted by probes f and g showed preferential expression in foetal and adult developmental stages[77].

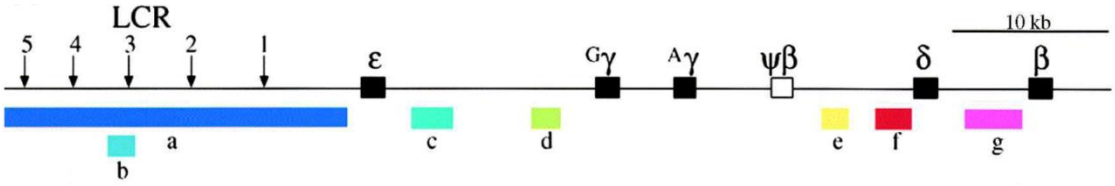


Figure 1- 14 Intergenic activity in the β-globin locus.

Vertical arrows indicate the positions of HSs 1–5 of the LCR. Genes are shown as black boxes, and the pseudogene (ψβ) as an open box. Coloured boxes (a–g) indicate the position of the restriction fragments used as probes for RNA FISH to detect intergenic transcripts. Adapted from: Gribnau, 2000, p.378[78].

D) Cell-cycle-specific timing of intergenic transcription

Intergenic transcription levels change over time as it is a cell-cycle- and locus-dependent mechanism. Gribnau *et al.* showed that intergenic transcripts are only detected in a proportion of cells in an unsynchronized population, suggesting that are generated in a cell-cycle-dependent manner; detectable predominantly during G1 phase and in a small percentage in early S-phase[77]. Miles *et al.* performed a high-resolution, locus-wide analysis of intergenic transcripts and histone modifications across the human β-globin locus during development in yeast artificial chromosome (YAC) transgenic mice and in primary human erythroid cells. They showed that the locus is composed of multiple chromatin sub-domains that are developmentally regulated, which can be distinguished by differential general sensitivity to DNase I, intergenic transcription and active histone modifications primarily to H3. They observed that the intergenic transcription pattern is complex and extensive and that active histone modifications strongly correlate with areas of non-S phase intergenic transcription linking the cell-cycle-specific timing of intergenic transcription with large chromatin domains of modified histones[50]. All these results support the link between intergenic transcription and the developmentally regulated chromatin remodelling across the β-globin locus.

E) Developmental changes in chromatin structure in the β-globin locus

The results of Gribnau *et al.* for intergenic transcription and hence for domain activation suggested that a progressive mechanism exists for opening and closing the β-globin locus domains during development (Figure 1-15). During embryonic and early foetal

stage, the LCR and εγ domains adopt an open configuration to enhance the expression of the ε- and γ-globin genes. As development continues, ε-globin gene is silenced and the δβ domain adopts now the accessible configuration. In the adult developmental stage γ-globin gene silencing is preceded by the adoption of a close configuration by the εγ domain[77].

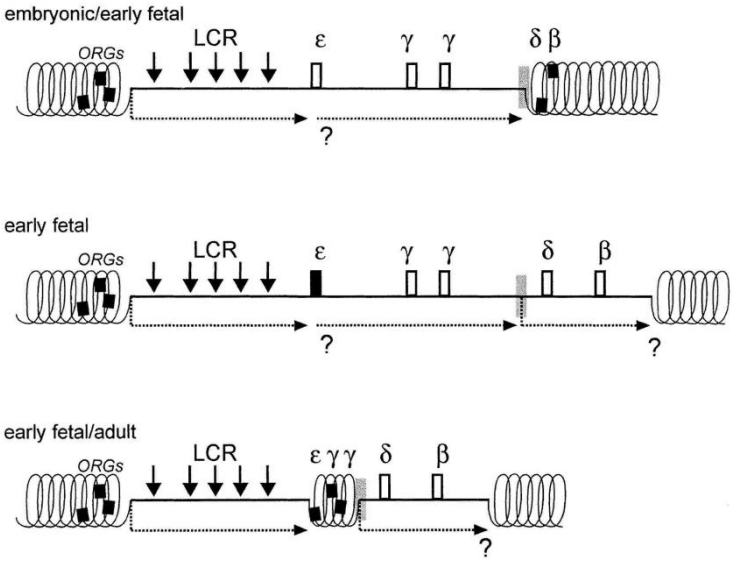


Figure 1- 15 Schematic Depiction of the Chromatin Structure Changes in the β-globin Locus during Development.

Shown are hypothetical, higher order chromatin structures to represent areas of high (straight line) and intermediate (coiled line) general sensitivity to DNase I. LCR HSs are shown as vertical arrows. Open boxes indicate active genes, whereas black boxes correspond to inactive genes. Shaded boxes indicate the region containing the putative chromatin boundary and δβ domain transcript initiation site. Areas delineated by intergenic transcripts are indicated by dashed arrows. Question marks stand for the initiation site of the εγ domain transcript and the 3′ end of the δβ domain transcript, which have not been establishedyet. Source: Gribnau, 2000, p.384[77].

F) Intergenic transcription and DNase sensitivity in the β-globin locus

The group of Gribnau *et al.* previously reported that intergenic regions in the β-globin locus are unidirectionally transcribed by RNAPII in a subset of erythroid cells in vivo[25]. These transcripts were predominantly detected in the LCR and adult β-globin genes and thus increased DNase I sensitivity. These results show a precise correlation between

chromatin structural domains and the extent of intergenic transcription, indicating that domains of active chromatin are delineated by intergenic transcription. They reported a correlation between histone acetylation, which is associated with modest changes in the regional DNase I sensitivity in the locus, and the domain-restricted intergenic transcription. Since histone acetyltransferases are associated with RNA–pol II complexes, intergenic transcription could play a role in decondensation of chromatin domains and gene activation. They also observed that a deletion of a 2.5 kb region between the γ- and δ-globin genes containing the putative adult sub-domain intergenic promoter results in a sub-domain-wide failure to adopt the accessible characteristic DNase I sensitive chromatin conformation during development, and an abnormally low and variegated expression of the adult β-globin gene. They concluded that intergenic transcription is an essential component of the remodelling process of chromatin subdomains to a hyperaccessible structures that determine which genes are able to interact with the LCR in erythroid cells[77].

G) Conflictive studies

The work of Plant *et al*.[16] appears to be contradictory to the suggested idea of Gribnau *et al*.[77] and Miles *et al*.[50] of a developmentally regulated intergenic transcription in the β-globin locus. Plant *et al*., according to their results, show TSA-induced intergenic transcription along the β-globin locus, reported that intergenic transcription can be found along the β-globin locus in a not necessarily developmentally regulated manner.

More recently, Haussecker and Proudfoot argued that the results of their work were not consistent with the activation-linked models based on the interaction between intergenic transcript abundance and chromatin conformation. Their data support that there is no correlation between intergenic transcription and chromatin activation in the β-globin locus. Instead, they found that in cells with Dicer knocked down, there was an upregulation of the intergernic transcription within the β-globin locus, suggesting the involvement of the RNA interference pathway in the regulation of the non-coding transcription within the locus. Thus, they proposed that chromatin activation at the locus correlates with Dicer knockdown. Hence, their conclusion was that in erythroid cells, in transcriptional-inactive regions of the β-globin locus, chromatin is silenced by default intergenic transcription[22].

1.3 RNA Interference and RNA Silencing

RNA interference (RNAi) is a natural pthway by which double-stranded RNA (dsRNA) induces gene silencing by targeting complementary mRNA for degradation in eukaryotes[79]. Like the discovery of splicing, the unearthing of RNAi has everyone's attention focused on posttranscriptional phenomena[80]. RNAi represents a breakthrough in understanding how genes are turned on and off in cells[81]. One of the reasons for the interest in RNA silencing in recent years has been the potential to use this mechanism to study gene function, validate candidate drug targets, and perhaps even treat disease[82].

RNAi represents an evolutionarily conserved phenomenon first described in plants in 1990 and has subsequently been found in most eukaryotes[83; 84]. RNA silencing mechanisms were first identified as a cellular RNA-based immune system designed to interfere with the uncontrolled production of aberrant RNA molecules, which could be interpreted by the body as an opportunistic viral infection. The RNAi mechanism is extremely specific in targeting these RNA molecules. In fact, they are identified for degradation by well-defined complementarity rules between the targeting small RNA and target messenger RNA[85]. Although the RNAi is also involved in the prevention of genomic instability caused by mobile genetic elements such as transposons and repetitive elements which produce dsRNA intermediates inside the cells, they also employ these pathways to regulate the expression of their own genes[82; 86; 87;88].

1.3.1 History

The first experimental observation of what later became understood as RNAi was observed in plants in 1990 from the work of Napoli *et al.* attempting to create more attractive petunia flowers by genetically engineering additional flower pigmentation genes into the plant genome[89]. Surprisingly, instead of producing plants with more colourful flowers, many had lost most or all pigmentation and became white. When they looked at the expression of genes involved in the natural pigmentation biosynthesis pathway, they found that the mRNA level of the gene that corresponded to the newly introduced pigment gene had been strongly reduced[90]. This phenomenon, now collectively called RNA silencing, was originally named co-suppression (also termed post-transcriptional gene silencing) and quelling, respectively[85]. The first major

breakthrough in understanding the RNAi silencing mechanism came from studies of worms conducted by Fire et al. in 1998[91]. They tested the phenotypic effects of RNA molecules injected into the nematode worm *C. elegans* and found dsRNA, but neither antisense nor sense RNA reduced target mRNA levels suggesting that RNA-silencing was specific for dsRNA homologue to the corresponding mRNA[86].

1.3.2 small RNAs

For decades small RNAs have been ignored as mere degradation products of bigger transcripts[92]. According to their origin or function, two main categories of small RNAs involved in RNAi have been defined: short interfering RNAs (siRNAs), and microRNAs (miRNAs), which differ primarily in the nature of their respective precursors (Figure 1-16). siRNAs are generated from perfectly base-paired dsRNA precursors. miRNAs are generated from single-stranded precursor transcripts that fold into imperfectly base-paired hairpin structures[82; 87]. siRNAs were originally identified as intermediates in the RNA interference pathway by exogenous dsRNA. siRNAs are 21–25 bp dsRNA with a phosphate group at both 5' ends, and hydroxyl groups and two-nucleotide overhangs at both 3' ends, all hallmarks of RNaseIII-mediated cleavage[93]. Although siRNAs are typically thought of as originating from exogenous precursors or triggers, there have been identified siRNAs generated from endogenous precursors in plants, fungi, and animals[90].

Figure 1- 16 Small RNA duplexes.
siRNA duplex shows a perfect match, whilst miRNA duplex has a mismatch or a bulge in the centre. Adapted from: Siomi, 2004, p.401[93].

Small RNAs are processed by two dsRNA-specific RNase-III-type endonucleases, termed Drosha and Dicer. Drosha is only required for the maturation of miRNA precursors, but not in the cleavage of long dsRNAs[87]. Dicer processes both long dsRNA

into siRNAs and cytoplasmic miRNA precursors (pre-miRNAs) into mature miRNAs (Figure 1-17) [88].

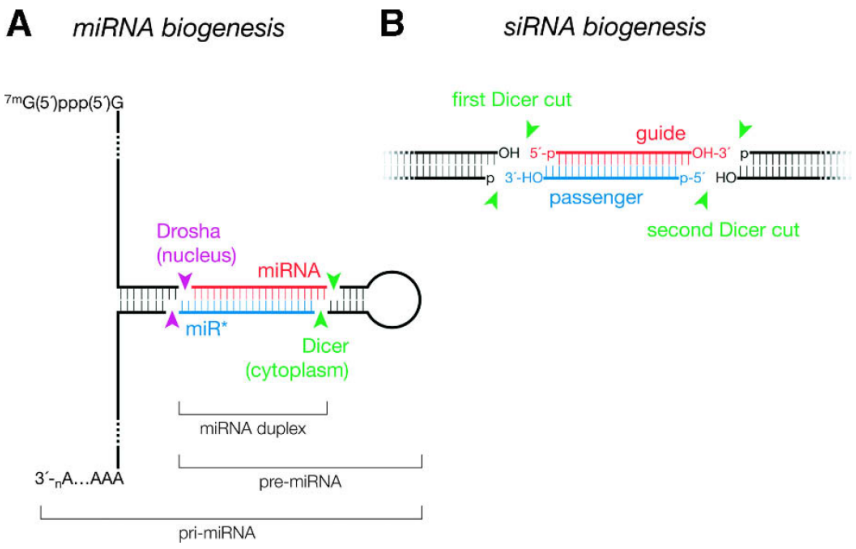


Figure 1- 17 Small RNA biogenesis in animals.

(A) miRNA are produced by the successive actions of two RNase III ribonucleases. After their transcription by RNA polymerase II, primary miRNA (pri-miRNA) are cleaved in the nucleus by Drosha. Drosha cleavage generates the pre-miRNA, which binds Exportin 5 and is exported to the cytoplasm. In the cytoplasm, Dicer is thought to bind the base of the pre-miRNA stem defined in the nucleus by Drosha. Dicer cleavage liberates a duplex comprising the miRNA and miRNA* strands of the pre-miRNA. (B) Long dsRNA is a substrate for Dicer, but not for Drosha. Dicer must make two successive pairs of cuts to yield an siRNA duplex. Dicer is thought to preferentially initiate dsRNA cleavage at the ends of dsRNA, a phenomenon that sometimes produces a phased string of siRNAs along the dsRNA. Source: Tomari, 2005, p.519[94].

1.3.3 Mechanism of RNAi

While the RNA gene silencing pathway is used by both siRNAs and miRNAs, there exist some important differences. The siRNA pathway begins with cytoplasmic cleavage of long double-stranded RNA by Dicer that result in short double-stranded RNA duplexes (siRNAs)[95]. siRNAs are then unwound, and only one strand is loaded into RISC. If the siRNA strand loaded into RISC has perfect sequence complementarity

with its target mRNA sequence, then site-specific mRNA cleavage takes place. This cleavage is a catalytic event, involving repeated cleavage of multiple copies of the target mRNA, leading to reduced copies of target mRNA and ultimately to inhibition of the target protein. In contrast to the siRNA pathway, the miRNA pathway begins with endogenously encoded primary miRNA transcripts in the cell nucleus that are processed into precursor microRNA and then exported from the nucleus into the cytoplasm. Here, Dicer further processes them for RISC loading. Unlike siRNAs, mature miRNAs typically have imperfect sequence complementarity to their RNA target sites which are typically limited to the 3' untranslated regions of the mRNA. Furthermore, interaction of a microRNA with its RNA target usually leads to direct translational inhibition rather than specific site-directed cleavage of the mRNA. Regardless of translational inhibition or cleavage, both microRNAs and siRNAs are able to regulate protein production at the RNA level in a very specific sequence-based manner[80; 85].

1.3.4 Assembly into RNA silencing effector complexes

Mature RISC, fully assembled with the guide strand of the siRNA direct mRNA cleavage, while RISCs loaded with a miRNA guide strand direct mRNA cleavage in plants but repression of translation in animals[88].

1.3.5 mRNA cleavage and translational repression

The target mRNA is cleaved by RISC whith the assistance of an RNA helicase in this process[87]. The products of the mRNA cleavage are further processed by cellular exonuclease activity[95] (Figure 1-18).

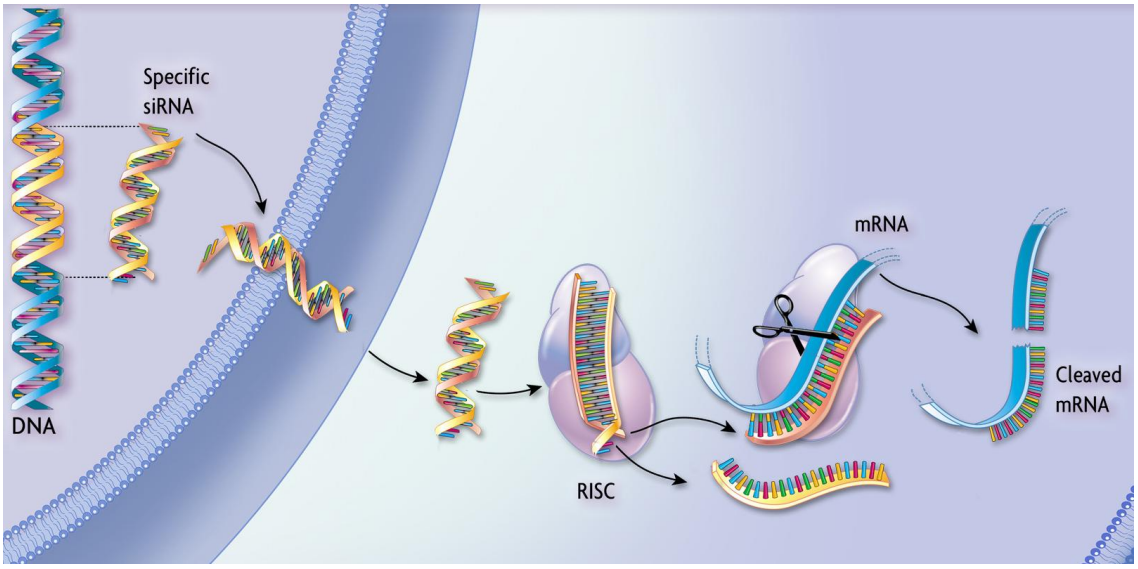


Figure 1- 18 The RNAi natural process.

Small interfering RNA (siRNA), a 21-25 base pair RNA strand, is targeted to a specific gene. Within cells, siRNA unwinds and is incorporated into RISC. siRNA is directed to a targeted messenger RNA (mRNA). The mRNA undergoes degradation, thereby interrupting the protein synthesis of the targeted gene. Source: Anylam, 2009, p.7[85].

1.3.6 RNAi techniques

It has been shown that the delivery of a dsRNA trigger can artificially induce the RNAi pathway. Long double-stranded RNAs (>200 nt) have been used to silence the expression of target genes in a variety of non-mammalian organisms and cell types. Introduced long dsRNAs trigger the RNAi pathway, being first processed into 20-25 nucleotide siRNAs by Dicer. Then, the siRNAs unwind and assemble into RISC. RISC is guided by the antisense siRNA strand to its complementary mRNA target, which will be eventually cleaved by RISC, in order to obtain specific gene silencing[79].

RNAi has been established today as a standard technique to investigate gene function or to identify cellular targets for specific intervention in many biological systems. RNAi can be activated at different stages by specific forms of regulatory RNAs, such as synthetic siRNAs, RNAs expressed as short hairpins RNAs (shRNA) and artificial miRNAs, efficiently targeting and causing degradation of the cognate RNAs[86]. The two most common methods used to activate RNAi are the transfection of cells with synthetic

siRNA duplexes, which then act to engage RISC. Intoduction into the cells of expression vectors that produce short haipin RNAs (shRNAs) which are then endogenously cleaved into siRNAs, represent an alternative to methods indicated above to trigger the RNAi pathway[83] (Figure 1-19).

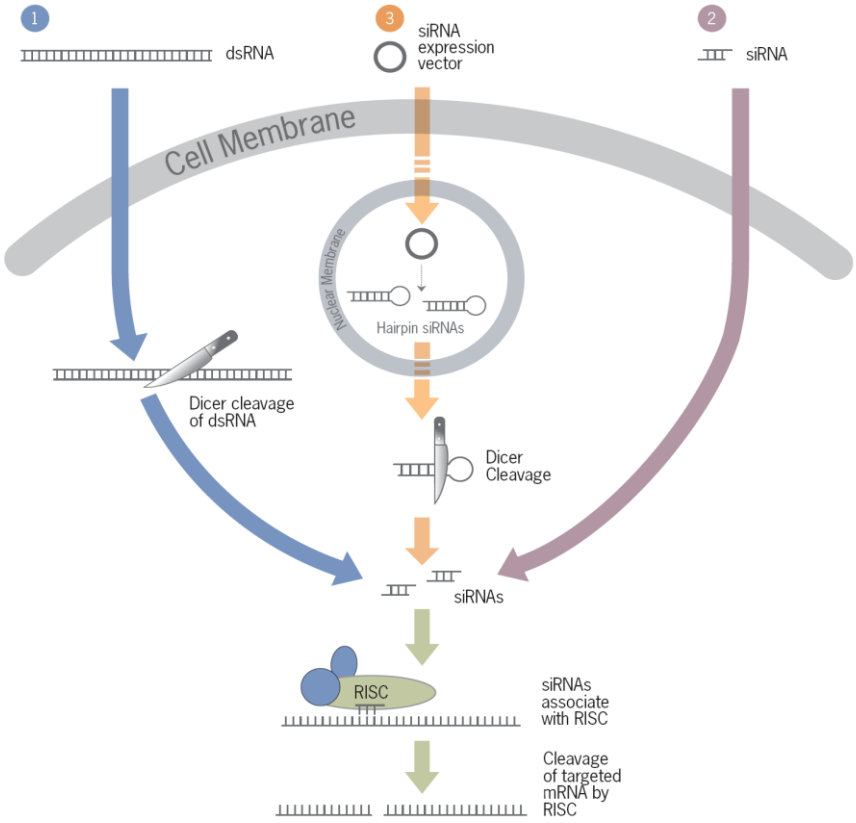


Figure 1- 19 Three Ways to Trigger the RNAi Pathway.

(1) In non-mammalian systems, the RNAi pathway commences when dsRNA (usually longer than 30 bp) is introduced into cells. In mammalian systems, RNAi can be triggered by siRNA molecules (2) or by DNA based expression vectors designed to express short hairpin RNA (shRNA) molecules (3). In each case, gene silencing results from destruction of mRNA that is complementary to the input siRNA (2) or the siRNA molecules created by Dicer cleavage of longer dsRNA (1) or shRNA (3) molecules. Source: Applied, 2009, p.3[79].

A) Synthetic siRNAs

Chemically synthesized or in vitro transcribed siRNAs are directly loaded into RISC after its transfection into the target cells. However, high intracellular siRNA concentrations (>100 nM) may induce non-specific effects due to saturation of the RNAi machinery, and the RNAi effect achieved is transient, lasting typically for 3–7 days[88]. In addition, the experimental phenotypes observed depend on the transduction rates, which may significantly differ between different cell types especially for primary cells[86].

B) Pol III promoter vector-based siRNA expression

Vector-based siRNA provide a low-cost alternative to the chemically synthesized siRNAs, which cost is high and their effect is transient[95]. In contrast to directly transfected siRNAs, shRNAs are endogenously transcribed, thanks to the development of vectors mediate transient or persistent gene silencing after stable viral-integration into the host genome[86]. The shRNA contains a perfectly double stranded stem of 19–29 bp with one strand identical in sequence to the target mRNA and 3′ U overhangs. The two strands of the stems are linked by a 9-bp long loop sequence, and the shRNA is processed by Dicer to generate functional active siRNA[86] (Figure 1-20). The eukaryotic H1 and U6 pol III promoters are the most commonly used to transcribe shRNAs. These promoters generate high levels of small, non-coding RNA transcripts lacking the polyA-tail and initiate from position +1 of the transcripts without any restriction by an inhibitory 5'-nucleotide. Furthermore, pol III promoters show high activity in all cell types and transcription terminates at a stretch of 4 thymidines, making it possible to produce short RNA with 1-4 uridines at the 3’ end[86].

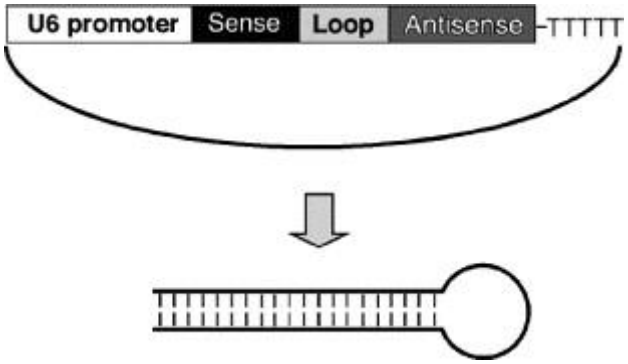


Figure 1- 20 Diagrammatic presentation of stem loop type siRNA driven by U6 vector.
Source: Wadhwa, 2004, p.80[95].

1.3.7 Mammalian RNAi Experiments

In mammalian cells, transfection of dsRNAs longer than 30 nt induced an antiviral response, which eventually leads to cell suicide. This obstacle was overcome in 2001 by Elbashir *et al*., who showed that synthetic versions of short double-stranded RNA molecules result in potent RNAi gene silencing in mammalian cells without inducing the interferon response[96]. These siRNA are long enough to induce RNAi, but small enough to avoid inducing an immune response[85]. For performing RNAi experiments in mammalian cell systems is necessary a gene specific siRNA or shRNA that effectively targets the gene of interest (RNAi trigger), an efficient delivery method for the siRNA or shRNA (generally transfection or electroporation), and robust transcriptomc and phenotypic assays to detect the RNAi effect, and proper controls both for siRNA delivery and for the phenotypic assay[88].

1.4 The k562 cell line

In 1975, Lozzio and Lozzio established the K562 cell line from leukemic cells obtained in December 1970 from a pleural effusion of a 53-year-old female suffering from chronic myeloid leukaemia (CML) for about 4 years, at the terminal stage of an acute blast crisis. The successful establishment of this cell line was due to the active proliferative capacity (K562 has a mean doubling time of 12 hr) of the highly undifferentiated cells obtained from the patient, which were already growing as a cell suspension in the pleural fluid and were not overgrown by the proliferation of normal

cells, as has occurred with most permanent cell-lines established from human leukemic blood[97]. K562 cells are thought to be very early precursors of erythroid cells. This cell line has also been widely employed in the investigation of alternative treatments for thalassemias and sickle cell disease[98].

The K562 cell line is composed of undifferentiated blast cell with a diameter of about 20 µm (Figure 1-21), and has properties in common with erythroid cells. In particular, cells of this line are rich in glycophorin A, which is found exclusively in human erythrocytes, and spectrin in their membranes, and may be induced to produce foetal and embryonic haemoglobin in the presence of 0.1 mM haemin[99].

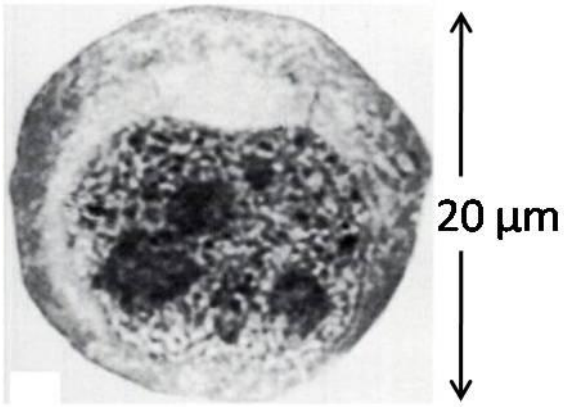


Figure 1- 21 Morphological characteristics of K562-undifferentiated blasts.
Adapted from: Koeffler, 1980, p.346[99].

The K562 cell has a basophilic cytoplasm containing no granules and there are two or more prominent nucleoli. The K562 cells have nearly 1.5 times the normal number of chromosomes: 70 XXX, - 13, - 17, +7, +9 (p11), plus small metacentric chromosome, Xp-, 3p-, 9p-, t(15;18) (q21;q23), r(22), or 22q-. Furthermore, they also have a small ring chromosome r(22) or retain the Philadelphia chromosome (22q-), demonstrating that this cell line is of CML origin[99] (Figure 1-22).

Figure 1- 22 Karyotype of K562 cells.

The modal number for K562 is between 66~72. Marker chromosomes are classified according to their centromeric origin and indicated with arrows. This cell only has two chromosome 11 homologues, while other cells have 3. There is cell-to-cell variation in the number of structurally normal chromosomes. Source: Gribble, 2000, p.4[100].

1.4.1 Induction of K562 cell line by haemin

Haemin is the ferric chloride salt of haem (ferric protoporphyrin IX), which is an iron-containing protoporphyrin that serves as the prosthetic group of the haemoglobin molecule. Haemin uptake by cultured erythroid cells down-modulates their surface receptors for tranferrin and for acid isoferritin, thus suggesting that the uptake of haemin is physiologically regulated. Uptake of haemin during the early stages of maturation may accelerate haemoglobin formation by affecting the rate of transcription of the globin genes, the processing and stability of their mRNA, or the efficiency of translation into globin chains. Externally supplied haemin is taken up and has a profound effect on proliferation and differentiation of haematopoietic progenitors, inducing a program of erythroid maturation in the K562 cell line, which includes the production of embryonic and foetal, but not adult, haemoglobin. Rutherford *et al*. investigated the capacity of K562 cells to differentiate *in vitro* and reported that they are induced by haemin to synthesise large amounts of haemoglobin. They found that the K562 cells showed a stable pattern of erythroid differentiation in response to haemin and that embryonic

haemoglobins were synthesized but adult haemoglobin synthesis was strictly repressed[101].

Fibach *et al.* reported that uptake of exogenously supplied haemin in erythroid cultures increased preferentially the production of foetal haemoglobin compared with adult haemoglobin; this was associated with a selective twofold elevation in y-mRNA levels. Thus, γ-globin levels in K562 cells can serve as a marker of induced cells. They suggested that the preferential increase in foetal haemoglobin is related to haemin's ability to accelerate early haemoglobin synthesis, involving more efficient processing of the primary transcripts. Haemin may also have a posttranslational effect, by providing heme for completion and stabilization of the haemoglobin molecules[102].

1.5 Project summary

1.5.1 The BGL3 non-coding transcript

Non-coding RNAs appear to play a role in gene regulation by modulating chromatin structure. RNA FISH using probes against the intron of a gene produces compact signals; even the most actively transcribed genes only give small signals. Recent evidence shows that some ncRNAs can aggregate in RNA-protein complexes and bind at the promoters of genes, producing a much more diffuse visualisation pattern in RNA FISH images. Previous unpublished work from the Fraser lab shows that the RNA FISH images of a short non-coding transcript in the globin locus (termed BGL3) are intense diffuse signals (Figure 1-23).

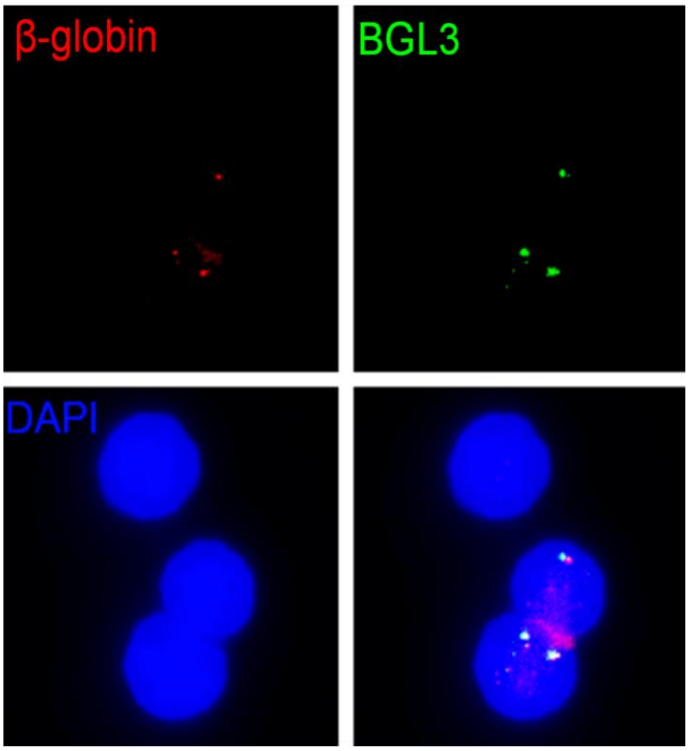


Figure 1- 23 Unpublished RNA FISH image from the Fraser lab at the Babraham Institute.
Probe against the intron of the β-globin gene produces compact signals. Probe against the BGL3 transcript produces intense diffuse signals.

This pattern of FISH signals suggests that the BGL3 transcript may be involved in an RNA-protein interaction and may be interacting with chromatin in the β-globin locus as part of a regulatory function within the locus (Figure 1-24).

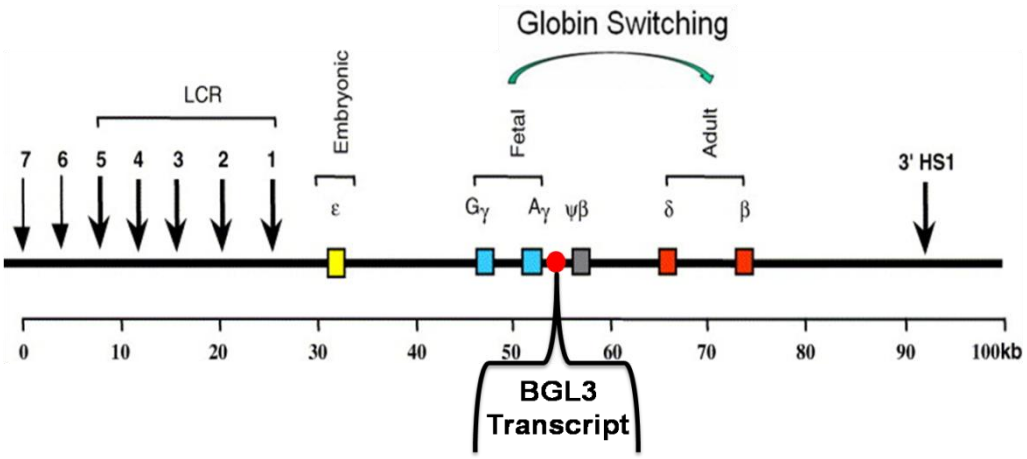


Figure 1- 24 Position of the BGL3 transcript on the β-globin locus.
Adapted from: Stamatoyannopoulos, 2005, p.263[10]

1.5.2 Aims and objectives

This project aims to test the role of BGL3 in the regulation of the human β-globin locus during development. The first step of the project will be to determine the expression profile of the BGL3 transcript and the γ- and β-globin genes, using qRT-PCR, in haemin-induced K562 cells, which have strikling similarities with erythrocytes and have been widely used for studying developmental expression across the β-globin locus.

The next step will be to test if the BGL3 transcript actually plays a role in regulating transcription within the β-globin locus. To this end, it will be attempted to perturb the levels of BGL3 using overexpression and knockdown plasmids, which will be transfected onto K562 cells. The effect of the overexpression and knockdown of the BGL3 transcript will be tested on the activity of the γ- and β-globin genes, which will be quantified using qRT-PCR. The results that will stem from this research project might be useful to obtain a deeper understanding of the role of intergenic transcription within the β-globin locus and its involvement on chromatin remodelling and in haemoglobin switching, providing a new tool for the development of treatments for sickle cell disease and β-thalassaemia.

2 Materials and methods

2.1 Bioinformatic tools

2.1.1 UCSC Genome Browser

The University of California, Santa Cruz, (UCSC) Genome Browser (available at http://genome.ucsc.edu/) was used to extract the genomic sequence of the BGL3 gene of the β-globin locus. A custom track was also generated so that designed primers and siRNAs could be viewed alongside the BGL3 gene.

2.1.2 BLAT

The BLAT tool (available at http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) was used to check for complementarities between the designed siRNAs and primers and the rest of the human genome.

2.1.3 Primer3

Primer3 is an open source, community-development program used to design PCR primers, hybridisation probes and sequencing primers (available at http://frodo.wi.mit.edu/primer3/). Primer3 was used to design paired primers and the hybridisation probe for the RT-PCR experiments on BGL3 expression.

2.1.4 NEBcutter

NEBcutter (available at http://tools.neb.com/NEBcutter2/index.php) was used to check that the pattern of restriction digestions used in the cloning procedure of the pEF6/V5-His A vector is compatible with the DNA sequence of the BGL3 gene.

2.1.5 RNAfold

The RNAfold web server (available at http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) was used to predict and visualize the secondary structure of the BGL3 ncRNA.

2.1.6 Translate Tool

The Translate Tool (available at http://www.expasy.ch/tools/dna.html) from the ExPASy web server was used to translate nucleotide sequence of the BGL3 transcript to

a protein sequence in order to determine the translation stop codons content of the BGL3 transcript.

2.2 Cell biological methods

During this project K562 cells were utilized, which were kindly provided by Peter Fraser (Babraham Institute, Cambridge). Manipulation of the cells, culture medium and plasticware was performed in a laminar flow cabinet precleaned with 70% isopropyl alcohol according to the standard cell biology procedures.

2.2.1 Materials

Component	**Supplier**	**Catalogue number**
RPMI medium 1640 (1X) +L-glutamine	Gibco	21875
Foetal bovine serum	Gibco	10270-106
Penicillin-Streptomycin	Gibco	15140
DMSO	Sigma-Aldrich	D-5879
Haemin	Fluka	51280
Isopropyl alcohol	Sigma-Aldrich	I9516

2.2.2 Medium preparation

The K562 cell line was maintained in logarithmic growth in RPMI medium 1640 (1X) +L-glutamine. The medium was supplemented with a suspension of 50 U/mL penicillin and streptomycin and 10% (v/v) of foetal bovine serum (FCS).

2.2.3 Cultivation of K562 cells

1. The cell suspension stored in liquid nitrogen was warmed up from -170°C to room temperature by dipping the vial in a 37°C water bath.
2. 1 mL of cell suspension (1 x 10^7 cells) was dilutssed in 9 mL of RPMI 1640 medium with supplements and centrifuged for 5 minutes at 1,000 x g. The supernatant was discarded.
3. The cell pellet was then resuspended in 2 mL of fresh medium and the cells were transferred into a T25 cell culture flask and incubated overnight at 37°C and 5% CO_2.
4. After 24 hours 10 mL of fresh medium were added and the cells were incubated for two more days.
5. Cells were split in a 1:5 ratio three times per week with a final culture volume of 40 mL in a T75 flask and incubated as before.

2.2.4 Induction of K562 cells with haemin

As described in the introduction, exogenously supplied haemin in K562 cultures preferentially increases the production of foetal haemoglobin compared with adult haemoglobin; this is associated with a selective elevation in γ-mRNA levels. The globin induction experiments were performed supplementing the media with a concentration of 100 μM of haemin.

2.2.5 Cryopreservation

1. A suspension of 40 mL of K562 cells, with a concentration of $1.25x10^5$ cells/mL, was centrifuged at 1,000 x g for 5 minutes and the pellet was subsequently resuspended in freezing medium, which is the medium used for cultivation supplemented with 5% DMSO.
2. Finally 1 mL of cell suspension with a concentration of $5x10^6$ cells/mL was transferred into a cryo vial, wrapped in tissue and stored in a box at -80°C over night.
3. The next day, cells were transferred into the liquid nitrogen tank and kept at -170°C.

2.3 Molecular biology methods

2.3.1 Haemoglobin quantification in K562 cells

Haemoglobin production in K562 was quantified by measuring haem content in cells.

A) Materials

Component	Supplier	Catalogue number
Igepal CA-630	Sigma-Aldrich	I8896-50ML
Tris 7.5	Fisher	BPE 152-1
NaCl	BDH	102415K
$MgCl_2$	BDH	101494V
Spectrophotometer	WPA	S2000

B) Protocol

1. 10X RSB buffer was prepared as follows:

Component	Concentration
Tris 7.5	100 mM
NaCl	100 mM
$MgCl_2$	30 mM
R.O. water	To a final volume of 1 L

2. 10^6 K562 cells were centrifuged for 5 minutes at 1,000 x g.
3. The supernatant was discarded and the pellet was resuspended in 1 mL of ice cold 0.5X RSB.

4. The cell suspension was incubated on ice for 5 minutes.
5. 50 µL of Igepal CA-630 was added to the suspension.
6. The cell suspension was biefly vortexed and centrifuged for 5 minutes at 1,000 x g.
7. Absorbance of the supernatant at 415 nm (visible) was measured using a spectrophotometer.
8. Absorbance of 1.0 at 415 nm is equivalent to 0.0945 mg/mL of haemoglobin.

2.3.2 RNA extraction from K562 cells

RNA extraction from K562 cells was performed with the RNeasy Mini Kit, QIAshredder and microcentrifuge.

A) Materials

Component	Supplier	Catalogue number
RNeasy Mini Kit	Qiagen	74104
QIAshredder	Qiagen	79654

B) Protocol

1. 5 x 10^6 K562 cells grown in suspension were pelleted by centrifuging for 5 minutes at 1,000 x g. The supernatant was discarded.
2. Cells were disrupted by adding 600 µl of Buffer RLT and vortexing them. This buffer contains guanidine-thiocyanate, which immediately inactivates RNases to ensure purification of intact RNA.
3. The lysate was transferred directly into a QIAshredder spin column placed in a 2 ml collection tube, which was the centrifuged for 2 minutes at full speed, in order to homogenise the cell lysate.
4. 600 µl of 70% ethanol was added to the homogenized lysate and mixed by pipetting to provide appropriate binding conditions.
5. The sample was then transferred to an RNeasy spin column placed in a 2 ml collection tube, which was centrifuged for 15 seconds at 8,000 x g. The flow-through was discarded.
6. 700 µl of Buffer RW1 was added to the RNeasy spin column, which was then centrifuged for 15 seconds at 8,000 x g to wash the spin column membrane. The flow-through was discarded.
7. 500 µl of Buffer RPE were added to the RNeasy spin column, which was then centrifuged for 15 seconds at 8,000 x g to wash the spin column membrane. The flow-through was discarded.
8. 500 µl of Buffer RPE were added to the RNeasy spin column, which was then centrifuged for 2 minutes at 8,000 x g to wash the spin column membrane. The flow-through was discarded.
9. The RNeasy spin column was placed in a new 2 ml collection tube and was centrifuged at full speed for 1 minute.
10. The RNeasy spin column was placed in a new 1.5 ml collection tube and 50 µl of RNase-free water were added directly to the spin column membrane. The column was centrifuged at 8,000 x g for 1 minute to elute the RNA.
11. Step 10 was repeated reusing the same collection tube.

12. Finally, the extracted RNA was stored at -20 °C until further treatment.

2.3.3 DNase treatment

DNase treatments were performed to obtain genomic DNA-free RNA.

A) Materials

Component	Composition	Supplier	Catalogue number
Amplification Grade DNase I	1 mL of 1 unit/µL of enzyme in 50% glycerol, 10 mM Tris- HCl, pH 7.5, 10 mM $CaCl_2$, 10 mM $MgCl_2$	Sigma-Aldrich	D 5307
10X Reaction Buffer	1 mL of 200 mM Tris-HCl, pH 8.3, 20 mM $MgCl_2$	Sigma-Aldrich	R 6273
Stop Solution	1 mL of 50 mM EDTA	Sigma-Aldrich	S 4809

B) Protocol

1. The DNase reactions were set up as follows in 0.2 mL tubes:

Component	Volume
RNA in water	8 µL
10X Reaction Buffer	1 µL
Amplification Grade DNase I (1 unit/µl)	1 µL

2. The tubes were mixed gently and incubated at room temperature for 15 minutes.
3. The DNase reactions were stopped by adding 1 µL of Stop Solution to bind calcium and magnesium ions and to inactivate the DNase I.
4. The tubes were incubated at 70 °C for 10 minutes to denature both the DNase I and the RNA.
5. The tubes were stored at -20°C until the reverse transcriptase reactions were performed.

2.3.4 Reverse Transcriptase reaction

Single-stranded cDNA from total, DNase-treated RNA was synthesized using the High Capacity RNA-to-cDNA Kit. This kit consolidates the components of the reverse transcription reaction into two tubes, 2x RT Buffer and 20x Enzyme mix, including RNase inhibitor. The reverse transcriptase (RT) products were further used in PCR and qPCR reactions.

A) Materials

Component	Supplier	Catalogue number
High Capacity RNA-to-cDNA Kit	Ambion	4387406

B) Protocol

1. The components of the kit were thawed on ice.
2. The RT reactions, including the negative control, were set up in 0.2 mL tubes as follows:

Component	Volume per reaction	
	+RT reaction	-RT reaction
2X RT Buffer	10 µL	10 µL
20X Enzyme Mix	1 µL	-
RNA sample	9 µL	9 µL
Nuclease-free water	-	1 µL

3. The tubes were centrifuged to spin down the contents and to eliminate air bubles.
4. The RT reactions were performed in a thermal cycler using the following program:

Step	Temperature	Time
1	37°C	60 minutes
2	95°C	5 minutes
3	4°C	∞

5. The cDNA was stored at -20°C until it was used for PCR or qPCR reactions.

2.3.5 Strand specific reverse transcriptase on the BGL3 gene

Two-step RT-PCR in which cDNA is generated in the first step with a specific primer was performed to assess the contribution of sense and antisense transcripts to the global expression of the BGL3 transcript.

A) Materials

Component	Supplier	Catalogue number
Cloned AMV First-Strand cDNA Synthesis Kit	Invitrogen	12328-032
Primers	Sigma Genosys	See section 2.3.5.B

B) Protocol

The primer pair used to synthesize the BGL3 cDNA and in the further PCR was:

Primer	**Sequence (5'->3')**
Forward	GATCAAGCCTGTGCCAGACA
Reverse	GAACATCTAATAGTGTTGGGG
Amplicon length (bp)	113

1. In a 0.2 mL tube were mixed:

Component	**Volume per reaction**
DNase-treated RNA in water	9 µL
10 µM Primer (Forward, reverse or both)	1 µL (0.5 µL of each in case of using both)
10 mM dNTP Mix	2 µL

2. RNA and primer were denatured by incubating at 65°C for 5 minutes and then placed on ice.
3. The master reaction mix was prepared on ice by combining:

Component	**Volume per reaction**
5x cDNA Synthesis Buffer	4 µL
0.1 M DTT	1 µL
RNaseOUT (40 U/µL)	1 µL
DEPC-treated water	1 µL
Cloned AMV RT (15 U/µL)	1 µL

4. 8 µL of the master reaction mix was pipetted into each reaction tube on ice.
5. The reaction tube was transferred into a thermal cycler and was incubated at 50°C for 60 minutes.
6. Reaction was terminated by incubating at 85°C for 5 minutes.
7. cDNA was stored at -20°C until used for amplification with PCR (second step).

2.3.6 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify the cDNA gene sequences to be cloned into the pEF6/V5-His A vector, as well as the cDNA sequences from the two-step RT-PCR experiment. After PCR the obtained products were assessed on an agarose gel.

A) Materials

Component	Concentration	Supplier	Catalogue Number
HIFI PCR Buffer	x10	Invitrogen	52045
$MgSO_4$	50 mM	Invitrogen	52044
2'deoxyguanosine 5' triphosphate	100mM	NEB	28406521V
2'deoxythymidine 5' triphosphate	100mM	NEB	28406531V
2'deoxyadenosine 5' triphosphate	100mM	NEB	28406501V
2'deoxycytidine 5' triphosphate	100mM	NEB	28406511V
Platinum Taq DNA Polymerase HIFI	500 rxn	Invitrogen	11304-029
Primers	Supplied lyophilised	Sigma Genosys	See section 3.5.2
Nuclease free H2O	-	Sigma-Aldrich	W4502

B) Protocol

1. 5X PCR Mix was prepared as follows and stored at -20°C until required:

Component	Volume
10X PCR buffer	250 µL
$MgSO_4$ 50 mM	100 µL
Nuclease-free water	130 µL
dNTPs	5 µL each

2. Master Mix was prepared as follows just before performing the PCR:

Component	Volume per reaction
5X PCR Mix	4 µL
cDNA template	1.5 µL
Nuclease-free water	12.4 µL
Taq DNA polymerase	0.1 µL

3. Primers synthesised by Sigma Genosys were supplied lyophilised. Primers were reconstituted with the correct volume of TE buffer (stated on the product information sheet) providing a 100mM stock solution.
4. Primers were subsequently diluted to a 10 µM concentration.
5. The following was added to each PCR tube:

Component	Volume
Forward primer (10 µM)	1 µL
Reverse primer (10 µM)	1 µL
Master Mix	18 µL

Reaction tubes containing RNase-free water instead of cDNA template were used as non-template control for the PCR reaction. The products of the –RT reactions were also used as a negative controls for the PCR reactions to show the effectiveness of the DNAse treatment.

6. The reaction tubes were then transferred to the thermal cycler for amplification of the cDNA template. The thermal cycler conditions were set up as follows:

Step	Temperature	Time	PCR Stage
1	95°C	5 minutes	Initial denaturation
2	95°C	1 minute	Denaturation
3	60°C	1 minute	Annealing
4	72°C	1 minute	Extension
5	Repeat steps 2-4	39 additional cycles	Amplification
6	72°C	5 minutes	Final extension
7	4 °C	∞	End of PCR

2.3.7 Real-Time Polymerase Chain Reaction

Relative quantification assays were used to analyze changes in BGL3 expression, as well as the γ- and β-globin genes of the β-globin locus, using Real-time Polymerase Chain Reaction (RT-PCR). RT-PCR was performed with a CFX96 Real-Time PCR Detection System coupled to a C1000 Thermal Cycler (Bio-Rad) using TaqMan chemistry. TaqMan Gene Expression Assays were performed using as a template single-stranded cDNA from total genomic DNA-free RNA samples. To quantify the normalized fold expression of the studied genes, three replicates were performed and all data were normalized using the endogenous GAPDH control.

A) Materials

Component	Supplier	Catalogue number
Inventiored TaqMan Gene Expression Assay	Applied Biosystems	4331182
Made-to-order TaqMan Gene Expression Assay	Applied Biosystems	4351372
TaqMan Gene Expression Master Mix	Applied Biosystems	4370048
C100 Thermal Cycler	Bio-Rad	184-1000
CFX96 Real-Time PCR Detection System	Bio-Rad	185-5096
Individual PCR tubes	Bio-Rad	TLS0801
Flat cap strips	Bio-Rad	TCS0803

TaqMan chemistry was used to study the expression levels of the BGL3, β-globin (TaqMan Gene Expression Assay Hs00747223_g1) and γ-globin (TaqMan Gene Expression Assay Hs00361131_g1) genes. The expression levels were normalized respect to the expression levels of endogenous GAPDH (TaqMan Gene Expression Assay Hs99999905_m1).

B) Protocol

1. The cDNA samples and TaqMan Gene Expression Assays were thawed on ice and resuspended by gentle vortexing.
2. The TaqMan Gene Expression Master Mix was gently mixed swirling the bottle.
3. The number of reactions was calculated taking in account that for each assay and cDNA sample three duplicates will be performed. No template controls (NTCs) reactions for each gene expression assay were also included.
4. The PCR reaction mix were prepared as follows:

PCR reaction mix component	**Volume per 20 µL reaction**	
	Single reaction	**Three replicates***
20X TaqMan Gene Expression Assay	1 µL	3.3 µL
2X TaqMan Gene Expression Master Mix	10 µL	33 µL
cDNA template	1 µL	3.3 µL
RNase-free water	8 µL	26.4 µL
* Replicate volumes include 10% excess for volume loss from pipetting.		

5. 20 µL of PCR reaction mix were transferred into each well of the reaction strips.
6. The wells were sealed with the cap strips and then loaded into the instrument.
7. A plate document was created and the RT-PCR was ran using the following thermal cycler parameter values:

Step	**Temperature**	**Time**
1	50°C	2 minutes
2	95°C	10 minutes
3	95°C	15 seconds
4	60°C	1 minute
5	Repeat steps 3-4	39 addtional cycles

8. Results were analyzed using the software of the CFX96 Real-Time PCR Detection System. Controls and samples were normalized for endogenous glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression when endogenous transcript levels were measured.

2.3.8 Transformation of competent K12 E. coli cells

Competent K12 *E.coli* cells were used to clone the plasmids pSilencer 2.1-U6 and pEF6/V5-His A. These plasmids contain an ampicillin resistance gene, therefore growth

of bacteria was regulated by inclusion of ampicillin into the LB agar once they had undergone the transformation procedure. All growth incubations were carried out in a shaking incubator at 37°C unless stated otherwise.

A) Materials

Component	Supplier	Catalogue number
LB Broth	Fisher	17328
Nutrient agar No2	Fluka	70116
Ampicillin	Boehringer	835 269
NaCl	BDH	102415K
$MgCl_2$	BDH	101494V
$CaCl_2$	Sigma-Aldrich	C-3881
KCl	BDH	295945C
Tryptone	LAB	MC5
Yeast extract	Oxoid	L21
Glucose	Sigma-Aldrich	G7528-1KG

B) Generation of competent K12 *E. coli* cells

1. Sterile LB broth was prepared (25 g/L) and sterilized at 121°C for 20 minutes.
2. 1 colony was picked off from an agar plate and aseptically transferred into a sterile 5 mL aliquot of LB broth and then incubated at 37°C overnight whilst been shaken (approximately 150 rpm).
3. The culture was placed on ice for 10 minutes.
4. Cells were harvested by centrifuging the 5mL culture at 1,600 x g for 5 minutes.
5. The supernatant was discarded and the cells re-suspended in 10 mL of sterile 0.1M cold $MgCl_2$.
6. The cells were spun down again at 1,600 x g for 5 minutes and re-suspended in 1ml of sterile cold $CaCl_2$ and then kept on ice for 2 hours.

C) Transformation of competent K12 *E. coli* cells

1. Super optimal broth with catabolite repression (SOC) medium was prepared and autoclaved as follows:

Component	Quantity
Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g
KCl	0.186 g
$MgCl_2$	0.952 g
Glucose	3.603 g
Water	To a final volume of 1 litre

2. 200µl of competent cells were added to 10µl of ligated plasmid. The mix was then incubated on ice for 30 minutes.

3. Cells were then heat shocked at 42°C for 2 minutes and placed immediately on ice for another 2 minutes before adding 400 µL of pre-warmed SOC medium.
4. The cells were then incubated for 1 hour at 37°C in a shaker incubator to allow the ampicillin resistance gene to become active.
5. The bacterial suspension was then aseptically transfered into a 50 mL tube containing 5 mL of LB broth with a concentration of 100 µg/mL of ampicillin. The bacterial cells were incubated overnight at 37°C in a shaker incubator
6. The bacterial suspension was diluted to 1/106. 200 µL of the dilution was plated onto 100 µg/mL ampicillin-containing LB agar plates.
7. Bacterial colonies were counted following overnight incubation at 37°C. Single colonies were picked and placed into a 50 mL tube containing 5 mL of LB broth with a concentration of 100 µg/mL of ampicillin. These were grown overnight at 37°C in a shaker incubator prior to plasmid extraction.
8. Glycerol (20%) stocks of the colonies were also prepared and kept at -80°C for long-term storage.

2.3.9 Plasmid extraction

Plasmid extractions from transformed bacterial cultures were performed using the QIAprep Spin Miniprep Kit and a Microcentrifuge.

A) Materials

Component	**Supplier**	**Catalogue number**
QIAprep Spin Miniprep Kit	Qiagen	27104

B) Protocol

1. A 5 mL overnight culture of transformed E.coli was centrifuged at 1,600 x g for 5 minutes and the supernatant was discarded.
2. LyseBlue reagent and 20 µL of RNase A provided within the kit were added to Buffer P1. Following mixing the solution was stored at 2-8°C.
3. Pelleted bacterial cells were resuspended in 250 µL of Buffer P1 and transferred to a microcentrifuge tube whilst ensuring there were no clumps of cells within the suspension.
4. 250 µL of Buffer P2 were then added and mixed by gently inverting 4-6 times until the solution became viscous and homogenously blue.
5. 350 µL of Buffer N3 were added and mixed immediately by inverting the tube 4–6 times until mixed until all trace of blue had gone and the suspension was colorless.
6. The suspension was then centrifugated for 10 minutes at 10,000 x g in a microcentrifuge.
7. The supernatant was applied to a QIAprep spin column by decanting and centrifuged for 30 seconds. The flow-through was discarded.
8. To remove trace nuclease activity the column was washed by adding 0.5 mL of Buffer PB and centriguging at 10,000 x g for 30 seconds. The flow-through was discarded.
9. Ethanol (96-100%) was added to buffer PE as directed.

10. The column was washed again adding 0.75 mL of Buffer PE and centrifuging for 30 seconds.
11. The flow-through was discarded and the column was centrifuged for 1 minute to remove residual wash buffer.
12. To elute plasmid DNA the spin column was placed in a clean 1.5 mL microcentrifuge tube. 50 µL of Buffer EB (10mM Tris-Cl, pH 8.5) was added to the column. The column was left to stand for 5 minutes and then centrifuged at 10,000 x g for 1 minute.
13. The eluted plasmid DNA was stored at -20°C.

2.3.10 Restriction enzyme digestion of plasmid DNA and PCR products

Restriction enzyme digestions were performed to linearise the extracted plasmids and produce cohesive ends for subsequent ligations. PCR products were also digested to produce cohesive ends compatible with those of the linearized plasmid. Restriction enzyme digestions were also used to demonstrate the presence of the insert in the pEF6/V5-HIS A plasmid construct.

A) Materials

Component	**Concentration**	**Supplier**	**Catalogue number**
*Bam*H I	20,000 units/mL	New England Biolabs	R0136S
Hind III	20,000 units/mL	New England Biolabs	R0104S
Xba I	20,000 units/mL	New England Biolabs	R0145S
Kpn I	10,000 units/mL	New England Biolabs	R0142S
BSA	10 mg/mL	New England Biolabs	B9001S
NEBuffer 2	10X	New England Biolabs	B7002S

B) Protocol

(i) Digestion of pSilencer 2.1-U6 vector

1. Digestion reaction was set up as follows in a 1.5 mL tube:

Component	**Volume per reaction**
Plasmid	23 µL
NEBuffer 2 (10X)	3 µL
BSA (10 mg/mL)	3 µL
*Bam*H I (20 U/mL)	1.5 µL
Hind III (20 U/mL)	1.5 µL

2. All components were kept on ice whilst preparing the reaction. The digestion was started by the addition of the enzymes.
3. The reaction mixture was incubated at 37°C for 90 minutes.
4. Following digestion enzymes were heat inactivated by incubating at 65°C for 20 minutes.
5. The digested plasmid was stored at -20°C following asses by gel electrophoresis of the digests (see section 2.3.11).

(ii) Digestion of pEF6/V5-HIS A vector

1. Digestion reaction was set up as follows in a 1.5 mL tube:

Component	**Volume per reaction**
Plasmid	23 µL
NEBuffer 2 (10X)	3 µL
BSA (10 mg/mL)	3 µL
*Bam*H I (20 U/mL)	1.5 µL
Xba I (20 U/mL)	1.5 µL

2. All components were kept on ice whilst preparing the reaction. The digestion was started by the addition of the enzymes.
3. The reaction mixture was incubated at 37°C for 90 minutes.
4. Following digestion enzymes were heat inactivated by incubating at 65°C for 20 minutes.
5. The digested plasmid was stored at -20°C following asses by gel electrophoresis of the digests (see section 2.3.11).

(iii) Digestion of PCR products to be ligated into pEF6/V5-HIS A vector

1. Digestion reaction was set up as follows in a 1.5 mL tube:

Component	**Volume per reaction**
PCR product	23 µL
NEBuffer 2 (10X)	3 µL
BSA (10 mg/mL)	3 µL
*Bam*H I (20 U/mL)	1.5 µL
Xba I (20 U/mL)	1.5 µL

2. All components were kept on ice whilst preparing the reaction. The digestion was started by the addition of the enzymes.
3. The reaction mixture was incubated at 37°C for 90 minutes.
4. Following digestion enzymes were heat inactivated by incubating at 65°C for 20 minutes.
5. The digested PCR product was stored at -20°C following asses by gel electrophoresis of the digests (see section 2.3.11).

(iv) Digestion of ligated pEF6/V5-HIS A plasmid construct to demonstrate the presence of the insert

1. Digestion reaction was set up as follows in a 1.5 mL tube:

Component	**Volume per reaction**
Plasmid	23 µL
NEBuffer 2 (10X)	3 µL
BSA (10 mg/mL)	3 µL
Kpn I (20 U/mL)	1.5 µL
Xba I (20 U/mL)	1.5 µL

2. All components were kept on ice whilst preparing the reaction. The digestion was started by the addition of the enzymes.
3. The reaction mixture was incubated at 37°C for 90 minutes.
4. Following digestion enzymes were heat inactivated by incubating at 65°C for 20 minutes.
5. The digested plasmid was assessed by gel electrophoresis of the digests (see section 2.3.11).

2.3.11 Gel electrophoresis of plasmid DNA and PCR products

A) Materials

Component	Supplier	Catalogue number
Electrophoresis Tank	Bio-Rad	Sub-Cell GT
GeneSnap	Syngene	Product version 7.07
MassRuler (x6) loading dye solution	Fermentas	R0621
MassRuler (ready to use) DNA ladder	Fermentas	#SM0403
Ethidium Bromide	Sigma-Aldrich	E7637
Agarose	Fisher	BP1356-100
Trisma Base	Fisher	BP152-1
Acetic acid	Fisher	A/0360/PB17
EDTA	Sigma-Aldrich	E7889

B) Protocol

1. 50X TAE buffer was prepared as follows:

Component	Quantity
Trisma base (2M final concentration)	242g
Acetic acid (1M final concentration)	57.1 mL
EDTA (50 mM final concentration)	200 mL
R.O. water	Volume was adjusted to 1 L

2. Agarose gels (0.8-2% w/v) were prepared by weighting agarose and mixing in 1X TAE buffer.
3. Agarose was dissolved by boiling the solution in a microwave.
4. When the gel was at 60°C approximately it was poured into the cast with the comb and allowed to set.
5. The gel was placed into the electrophoresis tank, which was then filled with sufficient 1X TAE buffer to cover the gel.
6. 10 µL of DNA ladder were added to the first lane
7. 1 volume of loading dye solution was mixed with 3 volumes of DNA sample to a maximum volume of 30 µL.
8. Gels were run at 200 V for 3 hours as a maximum depending on the agarose concentration and length of the gel.
9. Gels were stained placing them in an ethidium bromide solution (1µg/mL) for 30 minutes.
10. Gels were exposed to UV light and gel images were captured using the GeneSnap software.

2.3.12 Extraction of digested plasmids and PCR products form agarose gels

In preparation for ligation, linear plasmids and PCR products were extracted from agarose gels using the QIAquick Gel Extraction Kit and a microcentrifuge.

A) Materials

Component	Supplier	Catalogue number
QIAquick Gel Extraction Kit	Qiagen	28704

B) Protocol

1. DNA fragments were excised from a 2% (w/v) agarose gel with a clean, sharp scalpel trying to minimize the size of the gel slice by removing extra agarose.
2. Gel slices were weighed in a colourless tube. 3 volumes of QG buffer were added to 1 volume of gel (100 mg~100 µL). Buffer QG solubilizes the agarose gel slice and provides the appropriate conditions for binding of DNA to the silica membrane.
3. Samples were then incubated at 50°C until the gel had completely dissolved.
4. 1 gel volume of isopropanol was added to the sample and then mixed.
5. To bind DNA, the sample was applied to a QIAquick spin column and centrifuged for 1 minute at 10,000 x g. The flow-through was discarded.
6. 0.5 mL of Buffer QG was added to the column and centrifuged for 1 minute to remove all traces of agarose. The flow-through was discarded.
7. The column was then washed by the addition of 0.75 mL of Buffer PE was left to stand for 5 minutes. The column was then centrifuged at 10,000 x g for 1 minute.
8. Flow-through was discarded from the collection tube; the column was then centrifuged for an additional 1 minute at 10,000 x g in order to remove any residual ethanol.
9. The column was placed into a clean microcentrifuge tube. DNA was eluted by the addition of 30 µL of Buffer EB (10mM Tris-Cl, pH 8.5). The column was left to stand for 5 minutes followed by centrifugation for 1 minute at 10,000 x g.
10. Extracted DNA was stored at -20°C.

2.3.13 pEF6/V5-His A

A) Materials

Component	Supplier	Catalogue number
pEF6/V5-His A	Invitrogen	V961-20

B) Description

pEF6/V5-His A is a 5.8 kb vector designed for over-expression of recombinant proteins in mammalian cell lines (Figure 2-1). The vector contains the following elements:

1. Human elongation factor 1α-subunit promoter (hEF-1α) for high-level expression across a broad range of species and cell types.
2. Three reading frames to facilitate in-frame cloning with a C-terminal tag encoding the V5 epitope and a polyhistidine metal-binding peptide.
3. Blasticidin resistance gene (*bsd*) to allow for selection of stable cell lines using blasticidin.

4. Episomal replication in cell lines that are latently infected with SV40 or that express the SV40 large T antigen.

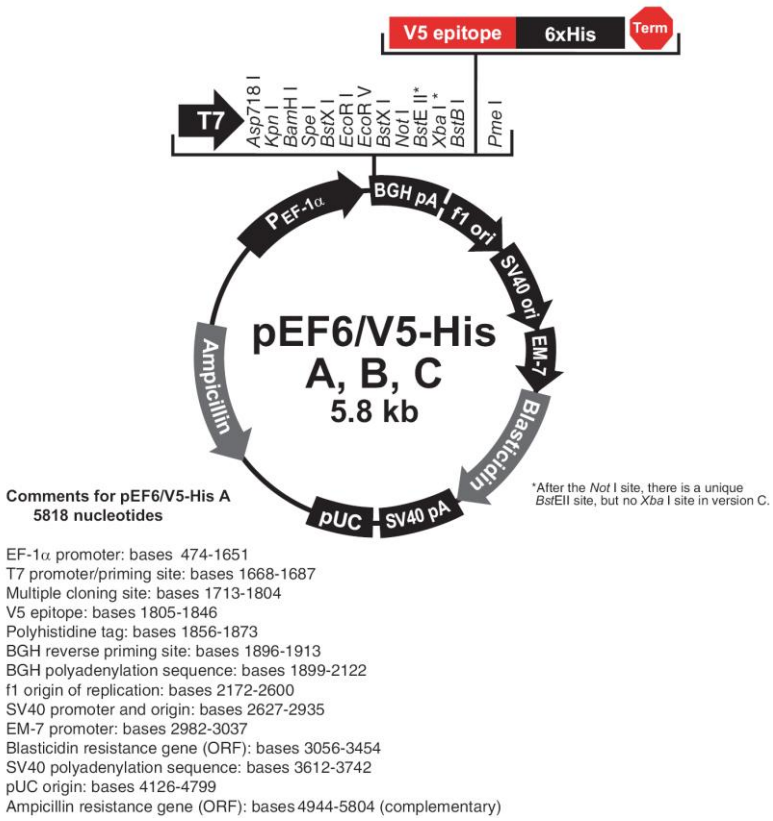


Figure 2- 1 Map of pEF6/V5-His.

This figure summarizes the features of the pEF6/V5-His vectors.

2.3.14 Blastidicin kill curve of K562 cells

Blastidicin in culture medium at 2-10 µg/ml is commonly used to select cells with an integrated plasmid containing the resistance gene. Cells without the resistance gene are normally killed within 3–5 days.

A) Materials

Component	Supplier	Catalogue number
Blastidicin	Sigma-Aldrich	15205-25MG
Trypan Blue	Sigma-Aldrich	T8154-20ML

B) Protocol

1. 20,000 K562 cells were plated into each well of a 24 well dish containing 1 mL of RPMI 1640 culture medium supplementd with FCS and antibiotics as described above.

2. After 24 hr, 500 µL of culture medium containing blastidicin concentrations of 10 µg/mL, 1 µg/mL, 100 ng/mL and 10 ng/mL were added into each of the 2 wells of each of the 4 sub group of wells of the dish. The rest of the wells were used as a control wells where were cultured without blastidicin.
3. Cells were cultured for 3 days and then the dishes were examined for viable cells. Cell viability was assessed each 24 hours using a haemocytometer and a solution of Trypan Blue 0.4% for vital staining.
4. The lowest blastidicin concentration that begins to give massive cell death in 3 days was used to select K562 cells containing the pEF6/V5-His A plasmid after transfection.

2.3.15 pEF6/V5-His A Ligation reactions

Digested PCR products were inserted into the pEF6/V5-His A vector performing the ligation reactions explained below.

A) Materials

Component	**Concentreation**	**Supplier**	**Catalogue number**
T4 DNA ligase	400,000 cohesive end units/ml	NEB	M0202L
Ligase buffer	10X	NEB	B202S
Nuclease-free water	-	Sigma-Aldrich	W4502

B) Protocol

1. Two 10 µL ligation reactions were set up; a plus-insert ligation, and the minus-insert negative control. To each tube, the following reagents were added:

Component	**Volume**	
	Plus-insert reaction	**Minus-insert reaction**
Digested and purified PCR product	1 µL	-
Nuclease-free water	-	1 µL
10X T4 DNA Ligase Buffer	1 µL	1 µL
Linear pEF6/V5-His A vector	7 µL	7 µL
T4 DNA ligase	1 µL	1 µL

2. The reactions were incubated for 3 hours at room temeperature and then the tubes were stored at –20°C for future transformation of *E. coli*.

2.3.16 pSilencer 2.1-U6 puro siRNA Expression Vector

The pSilencer 2.1-U6 puro siRNA Expression Vector was used to induce RNAi by the expression of siRNA within K562 cells.

A) Materials

Component	Supplier	Catalogue number
pSilencer 2.1-U6 puro siRNA Expression Vector	Ambion	AM5762

B) Description

The pSilencer 2.1-U6 puro siRNA Expression Vector features a human U6 RNA pol III promoter (Figure 2-2). The pSilencer 2.1-U6 puro vector contains the puromycin resistance gene *pac* from *Streptomyces alboniger*, which encodes puromycin N-acetyl-transferase and conveys resistance to puromycin to enable antibiotic selection in mammalian cells. The pSilencer 2.1-U6 puro vector also contains an ampicillin resistance gene, and an *E. coli* origin of replication. These two latter features are important to clone and amplify the plasmid in *E. Coli*. The plasmid can be linearized with both *Bam*H 1 and *Hin*d III to facilitate directional cloning.

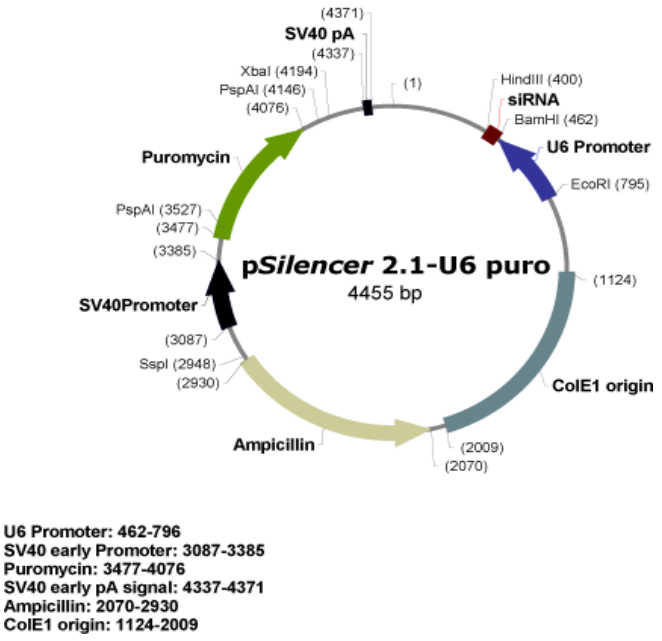


Figure 2- 2 pSilencer 2.1-U6 puro siRNA Expression Vector map.

2.3.17 Puromycin dihydrochloride kill curve of K562 cells

Puromycin was used to select cells carrying the pSilencer plasmid containing antibiotic the resistance gene.

A) Materials

Component	Supplier	Catalogue number
Puromycin dihydrochloride	Sigma-Aldrich	P8833-25MG
Trypan Blue	Sigma-Aldrich	T8154-20ML

B) Protocol

1. 20,000 K562 cells were plated into each well of a 24 well dish containing 1 mL of RPMI 1640 culture medium supplementd with FCS and antibiotics as described above.
2. After 24 hr, 500 µL of culture medium containing puromycin concentrations of 10 µg/mL, 1 µg/mL, 100 ng/mL and 10 ng/mL were added into each of the 2 wells of each of the 4 sub group of wells of the dish. The rest of the wells were used as a control wells where were cultured without puromycin.
3. Cells were cultured for 3 days and then the dishes were examined for viable cells. Cell viability was assessed each 24 hours using a haemocytometer and a solution of Trypan Blue 0.4% for vital staining.
4. The lowest puromycin concentration that begins to give massive cell death in 3 days was used to select K562 cells containing the pSilencer 2.1-U6 puro plasmid after transfection.

2.3.18 siRNA Template Design

Complementary 55–60 mer oligonucleotides with 5' single-stranded overhangs were designed for the target gene to be ligated into the pSilencer 2.1-U6 puro siRNA Expression Vector. The oligonucleotides encode 19-mer hairpin sequences specific to the mRNA target, a loop sequence separating the two complementary domains, and a polythymidine tract to terminate transcription.

A) siRNA Target Finder

Ambion's online siRNA Target finder (available at http://www.ambion.com/techlib/misc/siRNA_finder.html) was used to find potential sequences based on the design guidelines described in the reference Ambion, 2008[103]. The mRNA sequence was pasted into the window and the program generated a report indicating the potential siRNA target sites.

2.3.19 Hairpin siRNA Template Oligonucleotid Design

Below each siRNA target reported by the Ambion's online siRNA Target finder there was a link that sends the target directly to the web-based Insert Design tool for the pSilencer Vectors (available at http://www.ambion.com/techlib/misc/psilencer_converter.html). The hairpin oligonucleotides were designed according to the guidelines described in the reference Ambion, 2008[103].

2.3.20 Synthesis of hairpin siRNA template oligonucleotides for ligation into pSilencer 2.1-U6 puro siRNA Expression Vector

Oligonucelotides were synthesised by Sigma Genosys and were supplied lyophilised. Oligonucelotides were reconstituted with the correct volume of TE buffer (stated on product information sheet) providing a 100mM stock solution.

2.3.21 Cloning Hairpin siRNA Inserts into the pSilencer Vector

A) Materials

Component	Concentreation	Supplier	Catalogue number
T4 DNA ligase	400,000 cohesive end units/ml	NEB	M0202L
Ligase buffer	10X	NEB	B202S
Annealing solution	1X	Ambion	9613G2
Nuclease-free water	-	Sigma-Aldrich	W4502

B) Annealing of siRNA template oligonucleotides

1. A 1 μg/μL solution of each oligonucleotide in TE (10 mM Tris, 1 mM EDTA) was prepared
2. The 50 μL annealing mixture was assembled as follows:

Component	Amount
Sense siRNA template oligonucleotide	2 μL
Antisense siRNA template oligonucleotide	2 μL
1X DNA Annealing Solution	46 μL

3. The mixture was hetaed to 90°C for 3 min, and then placed in a 37°C incubator for 1 hour.
4. The annealed hairpin siRNA template insert was stored at –20°C for future ligation into the plasmid.

C) Ligation annealed siRNA template insert ino pSilencer 2.1-U6 Vector

1. 5 μL of the annealed hairpin siRNA template insert were diluted with 45 μL nuclease-free water for a final concentration of 8 ng/μL.

2. Two 10 µL ligation reactions were set up; a plus-insert ligation, and the minus-insert negative control. To each tube, the following reagents were added:

Component	**Volume**	
	Plus-insert reaction	**Minus-insert reaction**
Diluted siRNA insert	1 µL	-
1X DNA Annealing Solution	-	1 µL
10X T4 DNA Ligase Buffer	1 µL	1 µL
Linear pSilencer 2.1-U6 Vector	7 µL	7 µL
T4 DNA ligase	1 µL	1 µL

3. The reactions were incubated for 3 hours at room temperature and then the tubes were stored at –20°C for future transformation of *E. coli*.

2.3.22 Sequencing of plasmids

pSilencer plasmid constructs were sequenced to confirm that the inserts were correctly ligated into the plasmids and that there were no unwanted mutations. Purified plasmids were sent for sequencing at Geneservice Ltd. Plasmids were sequenced using the Forward T7 sequencing primer (5'-TAATACGACTCACTATAGGG-3'). Sequencing results were received by e-mail and analyzed using the FinchTV software (available at http://www.geospiza.com/Products/finchtv.shtml).

2.3.23 Transfection of K562 cells

Plasmids were transfected into the K562 cells using the electroporation method.

A) Materials

Component	**Supplier**	**Catalogue number**
RPMI medium 1640 (1X) +L-glutamine	Gibco	21875
Electroporator	Bio-Rad	Gene Pulser
Electroporator Cuvettes	Bio-Rad	1652088

B) Protocol

1. 10^7 K562 cells were centrifuged at 1,000 x g for 5 minutes and resuspended in RPMI 1640 medium without FCS or antibiotics to a final volume of 0.4 mL.
2. 100 µL of plasmid (10 µg approximately) were added to the cell suspension.
3. The mix was transferred to a sterile electroporation cuvette (0.4 cm of cuvette gap) and kept at room temperature for 15 minutes.

4. The electroporation cuvette was placed in the Gene Pulser Transfection Apparatus and the electroporation was performed under the following conditions:

Parameter	**Value**
Voltage	0.6 kV
Field Strength	1.5 V/cm
Capacitor	25 µF
Resistor	None
Time constant	0.6 milliseconds

5. After receiving the electric pulse, cells were centrifuged at 1,000 x g for 5 minutes and maintained pelleted for 20 minutes.
6. The electroporation medium was then removed and the pellet was resuspended in 10 mL of RPMI 1640 1X +L-glutamine medium supplemented with a suspension of 50 U/mL penicillin and streptomycin and 10% of FBS.
7. The culture was transferred to a T25 flask and was incubated for 24 hours at 37°C and 5% CO2.
8. The culture was transferred to a T75 flask containing 30 mL of culture medium and was incubated for 48 hours at 37°C and 5% CO2. The culture medium also contained of selective antibiotic (1 µg/mL of puromycin or 0.1 µg/mL of blastidicin) to enrich the culture for cells that were successfully transfected with the vector by killing off cells that lack the plasmid.
9. Cell viability was determined by Trypan blue exclusion method (see section 2.3.14).
10. 15 mL of the culture was used to extract RNA for further quantitative PCR analysis.
11. The remaining 15 mL culture was centrifuged at 1,000 x g for 5 minutes and resuspended in 30 mL of normal culture medium without selective antibiotic and was incubated at 37°C and 5% CO_2 for 7 days.
12. Cell viability was determined by Trypan blue exclusion method.
13. 15 mL of the culture was used to extract RNA for further quantitative PCR analysis.
14. The remaining 15 mL culture was splitted in two 7.5 mL cultures. 7.5 mL of culture media without selective antibiotic was added to each culture. One of the cultures was grown in the presence of haemin at concentration of 100 µM.
15. Cell cultures were incubated 84 hours at 37°C and 5% CO2. Cell viability was determined by Trypan blue exclusion method.
16. RNA extractions were performed for further quantitative PCR analysis.

3 Results

As described in the introduction, this project aims to test the role of BGL3 on the regulation of the human β-globin locus during development, as well as attempting to influence the expression levels of γ- and β-globin. First of all, the expression profile of the BGL3 transcript and the γ- and β-globin genes was analyzed in induced K562 cells using quantitative PCR. Then, two different approaches were taken to attempt to influence the expression of the BGL3 transcript in K562 cells. First, in order to over-express the BGL3 transcript, it was attempted to clone the BGL3 transcript in the pEF6/V5-His A plasmid, which is a constitutively expressed vector. Second, it was attempted to knockdown the BGL3 transcript using the pSilencer 2.1-U6 puro siRNA Expression Vector to induce RNAi pathway against the BGL3 transcript within the K562 cells. The main purpose of these experiments was to determine if there is a correlation between the expression levels of the BGL3 transcript and the γ- and β-globin genes of the human β-globin locus.

3.1 Characterization of the BGL3 transcript

The intense and diffuse signals yielded by the probes against the BGL3 transcript in the FISH images from the Fraser lab at the Babraham Institute (Figure 1-23) suggest that the transcript may be involved in an RNA-protein interaction and may be interacting with chromatin in the β-globin locus as part of a regulatory function within the locus. Thus, it is worthwhile to further investigate the expression profile of the BGL3 transcript to determine if it has a regulatory function within the β-globin locus.

3.1.1 Location of the BGL3 transcript within the β-globin locus

The UCSC genome browser (see section 2.1.1) was used to extract the sense and antisense sequence of the BGL3 gene (AY034471) from the human genome on chromosome 11, position chr11:5,222,361-5,223,352 (Figure 3-1). The UCSC genome browser displays results in a 5’ end to 3’ end design. However, the β-globin locus happens to be orientated in the opposite direction compared to how it is traditionally displayed (i.e. LCR on left, genes on right).

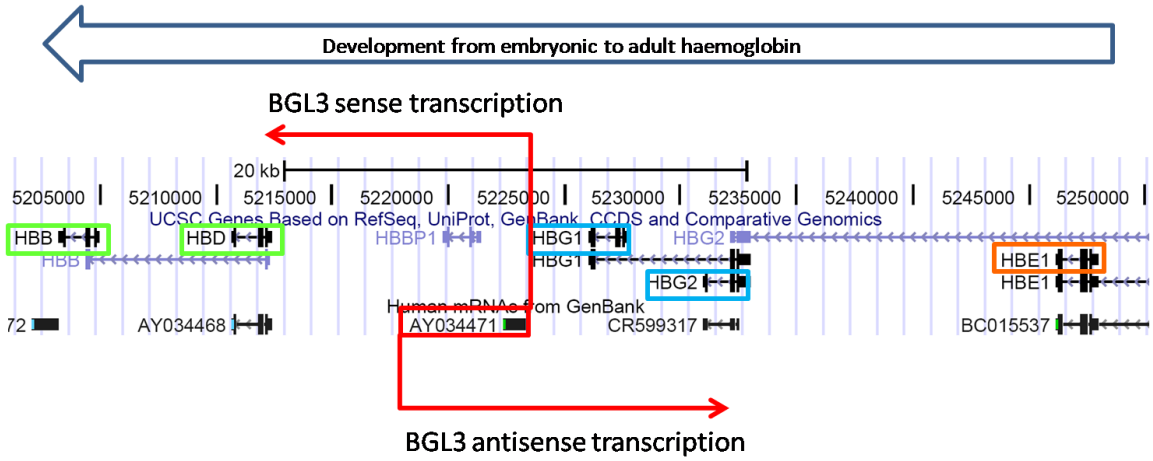


Figure 3- 1 BGL3 gene (AY034471) location on the β-globin locus.

Green rectangles: adult β-globin (HBB) and δ-globin genes; Blue rectangles: foetal $^{G}\gamma$- and $^{A}\gamma$- globin genes (HBG1 and HBG2); Orange rectangle: embryonic ε-globin gene (HBE1); Red rectangle: BGL3 gene (AY034471).

3.1.2 Coding potential of the BGL3 transcript

One of the ways of assessing the coding potential of the BGL3 transcript is translating its nucleotide sequence to its theoretical protein sequence. This was done using the Translate Tool from the ExPASy web server (see section 2.1.6). As can be seen in Figures 3-2 and 3-3, the protein sequences for the three reading frames of both sense and antisense BGL3 transcripts contain high numbers of translation stop codons, which prevent the translation of the transcript into protein and confirm its status of non-coding RNA.

Frame 1

LAVTAEFLVG **Stop** WKDGAAVHWYAGF **Stop** **Met** YVPKD **Met** RYGNEQKFFW
E **Stop** VLGPLKD **Met** T **Stop** SFL **Stop** GQSPQLNINVFPAYSQSCHFFFFISSISAL
KDNLGFASNCLSLQTFP **Met** FLSKTGNCLSTISSRLKYCVSNIRNSALIYWV
Stop LFEI **Stop** SLLEIH **Met** GVIS **Stop** TLFL **Stop** GFISGVTHKPAS **Stop** TSKYEDS
KP **Stop** EYKINCSSLCRWKCALSIPEIACLRR **Met** RRAAHETRAAACGILSVKN
VFA **Stop** LDIKTS **Stop** TQH **Stop** TIDQACARHK **Stop** PNAQHGPRNLLSSCLNR
AAHFSPNTIRCSGIIL **Stop** ICRI **Stop** HGLLFNDSEEISFVQISTLTTKWIKKHS
NKTQFSP **Stop** KKKKKKKK

Frame2

Stop Q **Stop** LLNSWLADGK **Met** GQLFTGTQGFRC **Met** YLRI **Stop** G **Met** A **Met** NR
NSFGNEF **Stop** GH **Stop** RT **Stop** PEVSSEASPHNSI **Stop** **Met** CFLHIVKVATSFSS
YHRSLLLKIILVLPQTVCHYKLSPCS **Stop** VKQVTASQLYQVD **Stop** NIVSLISEI
QL **Stop** YIGFNSLKFRVSLKYTWG **Stop** FPKLYSCKDLSQG **Stop** HTNQHPEPLS
Met RTVSLKNIK **Stop** TVLLSAGGSVPCLFLKLLV **Stop** DA **Stop** DVQH **Met** RHVQ
QPVEYCQ **Stop** R **Met** SLPD **Stop** I **Stop** RQVKHSIRL **Stop** IKPVPDTNDL **Met** PST
GHGISYPLA **Stop** TEQHTSPPTLLDVLA **Stop** FCRYVGFD **Met** DYCS **Met** IQRKS
PLFR **Stop** VH **Stop** LLNGLKNTVIKPSFPLKKKKKKKK

Frame 3

SSNC **Stop** IPGWL **Met** ERWGSCSLVRRVLDVCT **Stop** GYEVWQ **Stop** TEILLG
Met SFRAIKGHDLKFPLRPVPTTQYKCVSCI **Stop** SKLPLLFLHIIDLCS **Stop** R
Stop SWFCLKLFVTTNFPHVPK **Stop** NR **Stop** LPLNYIK **Stop** TKILCL **Stop** YQKFS
FNILGLTL **Stop** NLESP **Stop** NTHGGDFLNFILVRIYLRGNTQTSILNL **Stop** V
Stop GQ **Stop** ALRI **Stop** NKLFFSLPVEVCPVYS **Stop** NCLFETHETCST **Stop** DTCS
SLWNIVSEECLCLIRYKDKLNTALDYRSSLCQTQ **Met** T **Stop** CPARATESPILL
LEQSSTLLPQHY **Stop** **Met** FWHNFVD **Met** **Stop** DLTWTIVQ **Stop** FRGNLLCSDK
YTDY **Stop** **Met** D **Stop** KTQ **Stop** **Stop** NPVFPLKKKK

Figure 3- 2 Translated sequences corresponding to three reading frames of the sense BGL3 sequence.

Frame 1

F F F F F F F F F K G K T G F Y Y C V F **Stop** S I **Stop Stop** S V Y L S E Q R R F P L N H **Stop** T I V H V
K S Y I S T K L C Q N I **Stop Stop** C W G R S V L L C S S K R I G D S V A R A G H **Stop** V I C V W H R L
D L **Stop** S N A V F N L S L Y L I R Q R H S S L T I F H R L L H V S H V L H V S C V S N K Q F Q E **Stop**
T G H T S T G R E K N S L F Y I L K A Y C P H T **Stop** R F R **Met** L V C V L P L R **Stop** I L T R I K F R K S
P P C V F Q G D S K F Q R V K P N I L K L N F **Stop** Y **Stop** R H N I L V Y L I **Stop** L R G S Y L F Y L G T
W G K F V V T N S L R Q N Q D Y L **Stop** E Q R S **Met** I **Stop** R K R S G N F D Y **Met** Q E T H L Y **Stop**
V V G T G L R G N F R S C P L **Met** A L K L I P K R I S V H C H T S Y P **Stop** V H T S K T L R T S E Q L P
H L S I S Q P G I Q Q L L L

Frame2

F F F F F F F F L R G K L G F I T V F F N P F S S Q C T Y L N K G D F L **Stop** I I E Q **Stop** S **Met** S N P T
Y L Q N Y A R T S N S V G G E V C C S V Q A R G **Stop** E I P W P V L G I R S F V S G T G L I Y S L **Met**
L C L T C L Y I **Stop** S G K D I L H **Stop** Q Y S T G C C T C L **Met** C C T S H A S Q T S N F R N R Q G T L
P P A E R R T V Y F I F L R L T V L I L R G S G C W F V C Y P **Stop** D K S L Q E **Stop** S L G N H P H V Y
F K E T L N F K E L N P I Y **Stop** S **Stop** I S D I R D T I F **Stop** S T **Stop** Y S **Stop** E A V T C F T **Stop** E
H G E S L **Stop Stop** Q T V **Stop** G K T K I I F K S R D R **Stop** Y E E K E V A T L T I C R K H I Y I E L W
G L A S E E T S G H V L **Stop** W P **Stop** N S F P K E F L F I A I P H I L R Y I H L K P C V P V N S C P I F
P S A N Q E F S S Y C **Stop**

Frame 3

F F F F F F F F **Stop** G E N W V L L L C F L I H L V V S V L I **Stop** T K E I S S E S L N N S P C Q I L H I Y
K I **Met** P E H L I V L G E K C A A L F K Q E D R R F R G P C W A L G H L C L A Q A **Stop** S I V **Stop** C
C V **Stop** L V F I S N Q A K T F F T D N I P Q A A A R V S C A A R L **Met** R L K Q A I S G I D R A H F H
R Q R E E Q F I L Y S **Stop** G L L S S Y L E V Q D A G L C V T P E I N P Y K N K V **Stop** E I T P **Met** C I S
R R L **Stop** I S K S **Stop** T Q Y I K A E F L I L E T Q Y F S L L D I V E R Q L P V L L R N **Met** G K V C S D
K Q F E A K P R L S L R A E I D D **Met** K K K K W Q L **Stop** L Y A G N T F I L S C G D W P Q R K L Q V
Met S F N G P K T H S Q K N F C S L P Y L I S L G T Y I **Stop** N P A Y Q **Stop** T A A P S F H Q P T R N
S A V T A

Figure 3- 3 Translated sequences corresponding to three reading frames of the antisense BGL3 sequence.

3.1.3 Secondary structure of the BGL3 transcript

The secondary structures of the sense and antisense BGL3 transcripts were predicted and visualized using the RNAfold software (see section 2.1.5) (Figures 3-4 and 3-5). Both sense and antisense transcripts have complex secondary structures, which could be involved in their mechanism of action, and may affect the effectiveness of the designed siRNAs (see below) to target the sense and antisense BGL3 RNAs.

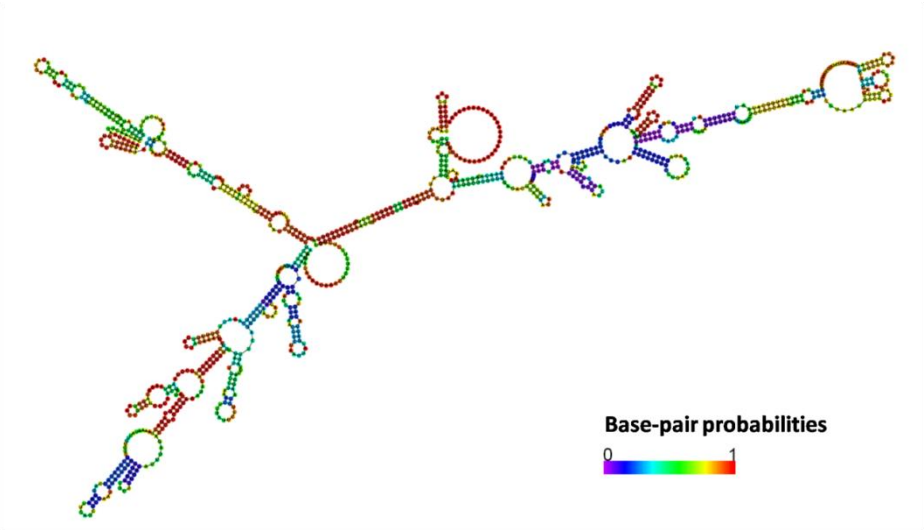


Figure 3- 4 Predicted secondary structure of the sense BGL3 transcript using the RNAfold software.

Bases are coloured according to the pair probability.

The optimal sense BGL3 transcript secondary structure corresponds to a minimum free energy of -255.60 kcal/mol. The free energy of the thermodynamic ensemble is -278.80 kcal/mol.

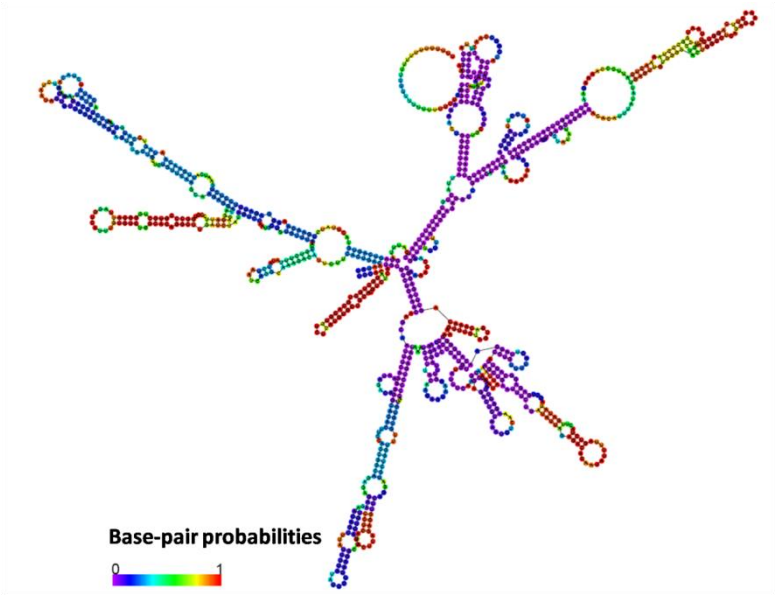


Figure 3- 5 Predicted secondary structure of the antisense BGL3 transcript using the RNAfold software.

Bases are coloured according to the pair probability.

The optimal sense BGL3 transcript secondary structure corresponds to a minimum free energy of -278.45 kcal/mol. The free energy of the thermodynamic ensemble is -299.85 kcal/mol.

3.2 Induction of K562 cells with haemin

Externally supplied haemin induces a program of erythroid maturation in the K562 cell line, which includes the production of large amounts of embryonic and foetal, but not adult, haemoglobin; this is associated with a selective elevation in γ-globin mRNA levels. Thus, elevated γ-globin levels in K562 cells can serve as a marker of induced cells (see section 1.4.1).

To check the reproducibility of the K562 cell line in our hands as a model of study to investigate intergenic transcription in the β-globin locus, as well as to determine the expression profile of the BGL3 gene, it was first tested if it was possible to induce our culture of K562 cell with haemin. As markers of induction the haemoglobin production levels and the γ-globin expression levels were compared between induced and non-induced K562 cells.

3.2.1 Experimental design

The main objective of this experiment was to compare different time course data from haemoglobin quantification and quantitative PCR experiments after induction of K562 cells. A culture of K562 cells was expanded and used to seed 28 flasks at a concentration of $2x10^4$ cell/mL (total volume 30 mL/flask). Cells were grown for 24 hours and split into two groups of 14 flasks. Haemin was added to one group of flasks at a final concentration of 100 μM. The cells in each flask were allowed to grow for a different amount of time (at 0 hours and every 24 hours until 144 hours) before haemoglobin production was quantified or RNA was extracted. Thus for each time point there are two biological replicates for both induced and non-induced K562 cells.

3.2.2 Haemoglobin production in K562 cells

Production of haemoglobin in K562 cells was quantified measuring haem (absorbs at 415 nm) content using a spectrophotometer (see section 2.3.1). As shown in Figure 3-6, there is a significant increase in the production of haemoglobin in haemin-induced compared to non-induced K562 cell cultures. The overproduction of haemoglobin

begins to be evident after 96 hours of culture in presence of haemin. The increase in haemoglobin production in non-induced K562 cells could be attributed to a fraction of cells that naturally stop proliferation and undergo erythroid differentiation even in the absence of haemin in the culture medium. Externally supplied haemin has been proved to induce the overproduction of haemoglobin (presumably embryonic and foetal, but not adult, haemoglobin) in our culture of K562 cells, indicating that the erythroid differentiation program has been triggered within the cells.

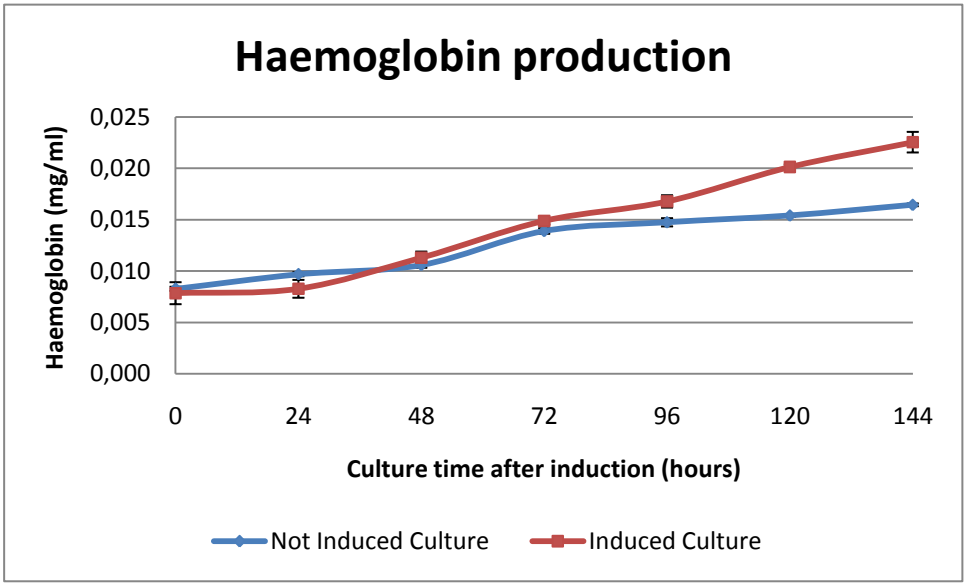


Figure 3- 6 Haemoglobin production in K562 cells.

Data obtained from two biological replicates for each time point with or without addition of haemin to the K562 cell cultures. Error bars represent standard deviation.

3.2.3 β-globin locus expression profile in K562 cells

The β-globin locus expression profile of induced and non-induced K562 cells was analyzed using quantitative RT-PCR. RNA form K562 cells was extracted using the RNeasy Mini Kit (Qiagen) (see section 2.3.2). Genomic DNA-free RNA was obtained with a DNase I treatment (Sigma-aldrich) and cDNA was synthesised using the High Capacity RNA-to-cDNA Kit (Ambion) (see section 2.3.4). TaqMan chemistry was used to study the expression levels of the BGL3, β-globin and γ-globin genes (see section 2.3.7). The expression levels were normalized respect to the expression levels of endogenous GAPDH. Normalized fold expression of the genes mentioned above in

induced K562 cells was calculated with respect to non-induced control K562 cultures grown in equivalent conditions. Results were analyzed using the software of the CFX96 Real-Time PCR Detection System. No genomic DNA was detected in the controls without the reverse transcriptase. PCR controls using water instead of cDNA as a template demonstrated no contamination.

A) Custom design of TaqMan Gene Expression Assay for the BGL3 gene

Since the BGL3 gene has no inventoried TaqMan Gene Expression Assay, the Primer3 software (see section 2.1.3) was used to design the primers (forward 5'- AGTGTTGG GGGAGAAGTGTG-3' and reverse 5'- TAGATCAAGCCTGTGCCAGA-3') and FAM probe (5'-ATAGGAGATTCCGTGGCCCGTG-3') for the custom-made BGL3 TaqMan Gene Expression Assay.

B) β-globin gene expression profile

As shown in Figure 3-7, the expression level of the β-globin gene in induced K562 cells does not show significant variations during the 144 hours of culture after addition of haemin to the culture medium.

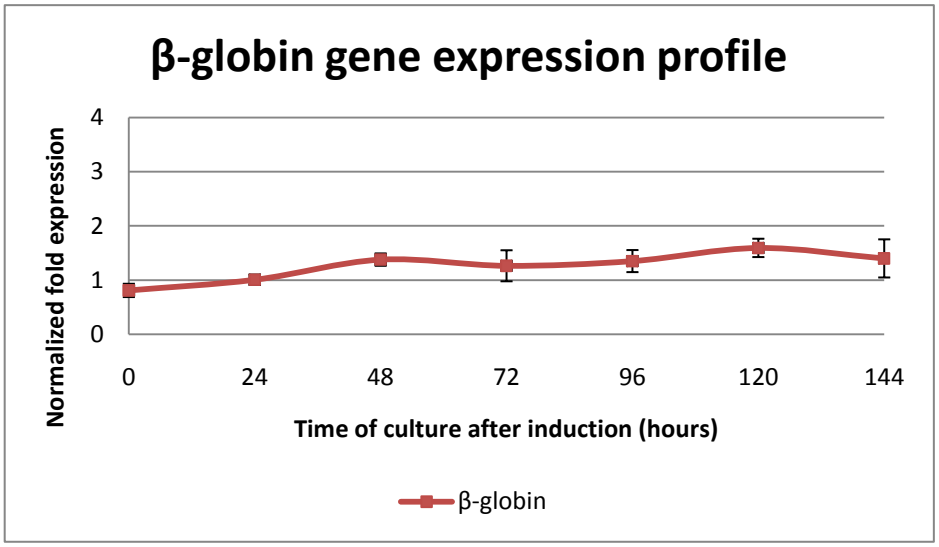


Figure 3- 7 β-globin gene expression profile of K562 induced cells by quantitative PCR.
Data obtained from three technical replicates of two biological replicates for each time point after addition of haemin to the induced K562 cell cultures normalised to non-induced cells. Error bars represent standard error of the mean.

C) γ-globin gene expression profile

As shown in Figure 3-8, the expression level of the γ-globin gene in induced K562 cells shows a significant increase of up to 3 fold 96 hours after addition of haemin to the culture medium.

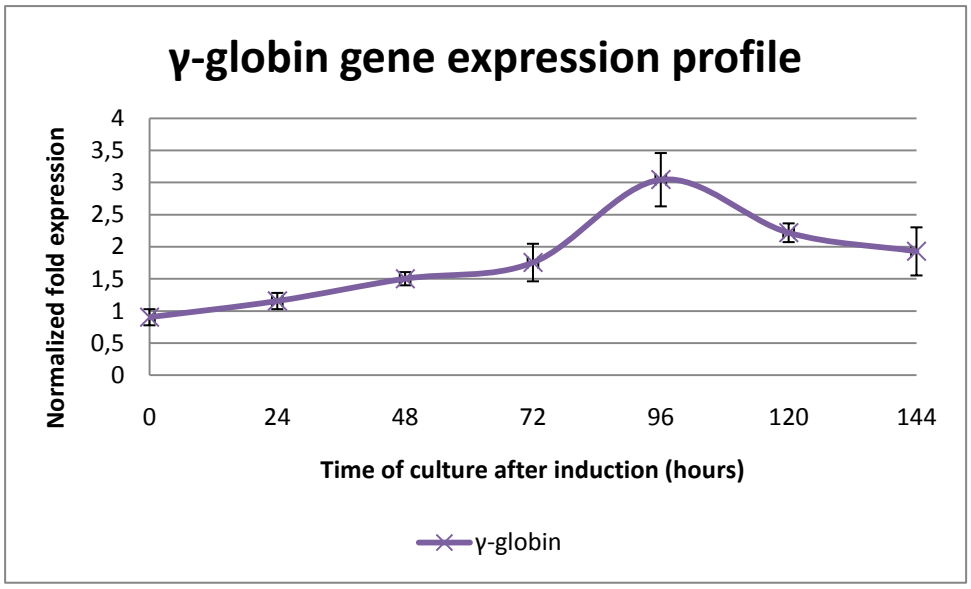


Figure 3- 8 γ-globin gene expression profile of K562 induced cells by quantitative PCR.

Data obtained from three technical replicates of two biological replicates for each time point after addition of haemin to the induced K562 cell cultures normalised to non-induced cells. Error bars represent standard error of the mean.

D) BGL3 gene expression profile

As shown in Figure 3-9, there is a significant increase of up to 3 fold in the expression level of the BGL3 gene 96 hours after addition of haemin to the culture medium.

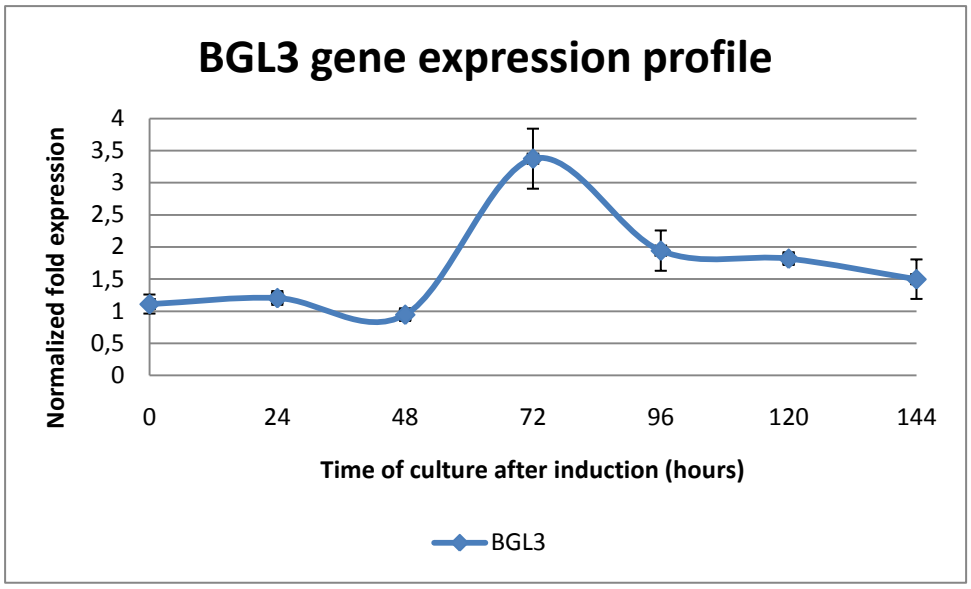


Figure 3- 9 BGL3 gene expression profile of K562 induced by quantitative PCR.

Data obtained from three technical replicates of two biological replicates for each time point after addition of haemin to the induced K562 cell cultures normalised to non-induced cells. Error bars represent standard error of the mean.

E) β-globin locus gene expression profile

Plotting together all our data referent to the expression levels of the BGL3, β-and γ-globin genes, interesting things can be observed (Figure 3-10).

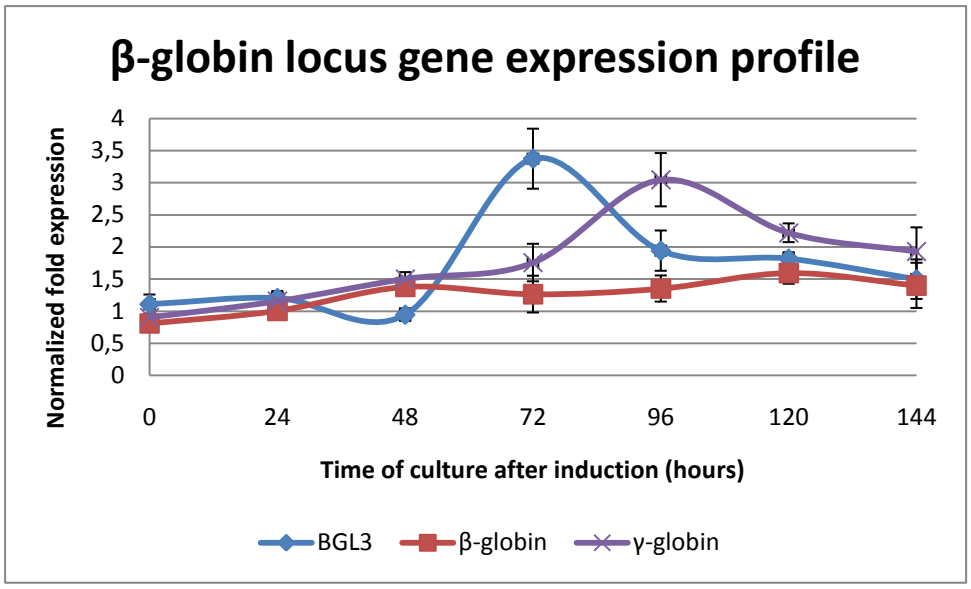


Figure 3- 10 β-globin locus gene expression profile of K562 induced by quantitative PCR.
Data obtained from three technical replicates per target of two biological replicates for each time point after addition of haemin to the induced K562 cell cultures normalised to non-induced cells. Error bars represent standard error of the mean.

As it was expected, due to the nature of the K562 cells, haemin-induced K562 cells undergo a program of erythroid maturation associated with a selective elevation in γ-globin mRNA levels of up to three fold. No significant changes can be observed in the expression levels of the β-globin gene. However, the big surprise comes from the expression profile of the BGL3 transcript in haemin-induced K562 cells. The data clearly shows that the BGL3 expression levels are actively regulated during differentiation of K562 cells into erythrocytes, suggesting a possible role of this non-coding transcript in the maturation program of the cells. Furthermore, the BGL3 transcript upregulation occurs just before the γ-globin gene reaches the peak of its expression. This suggests that the BGL3 transcript could be involved in the regulation of the γ-globin gene. Another possible explanation could be that the BGL3 transcript is associated with the repression of the β-globin gene during the K562 cells erythroid differentiation.

All these results, in addition to the haemoglobin quantification results, suggest that our K562 cells undergo erythroid differentiation upon haemin induction and that can be

used as a model to study intergenic transcription within the human β-globin locus. Moreover, the BGL3 expression levels have been shown to be actively regulated during erythroid differentiation of K562 cells, warranting further study of this non-coding transcript.

3.3 Direction of BGL3 transcription

The annotation of the BGL3 transcript suggests it is expressed in the same direction as the coding genes in the β-globin locus. As noted in Figure 3-1 this is termed as the sense transcript. However, it is known that some ncRNAs are transcribed in both directions, it is therefore important to determine the direction of BGL3 in K562 cells. By performing two-step RT-PCR (see section 2.3.5) in which cDNA is generated in the first step with a specific primer it is possible to assess the contribution of sense and antisense transcripts to the perceived levels relative to GAPDH shown in Figure 3-10. As shown in figure 3-11 the level of BGL3 product obtained when the reverse primer is used (which amplified the sense transcript) is greater than that obtained with the forward primer (which amplifies the antisense transcript). This suggests that the majority of BGL3 transcript is produced in the sense direction (as shown on Figure 3-1). However, some antisense product is still detected and it is therefore important in subsequent experiments to assess the effect of perturbing both sense and antisense transcripts.

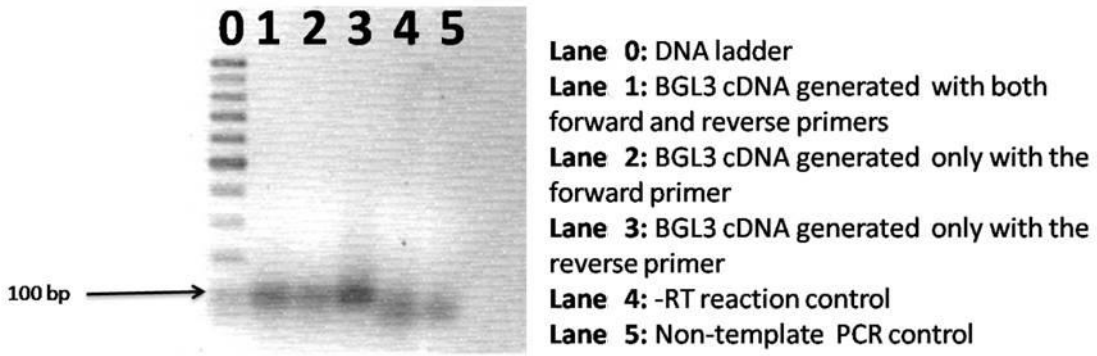


Figure 3- 11 Direction of BGL3 transcription

7 µL of PCR products were run on a 1.2% agarose gel for 45 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). Negative controls confirmed the absence of contamination or residual genomic DNA. PCR reactions produced one clear band of approximately 100 bp in size.

3.4 Generation of pSilencer 2.1-U6 puro plasmid constructs

To test wether the BGL3 transcript plays a direct role in regulating transcription in the human β-globin locus it was attempted to perturb the levels of BGL3 by knocking it down using the RNAi pathway. Vector-based siRNA provides a low-cost alternative to the chemically synthesized siRNAs, whose cost is high and effect is transient. In contrast to directly transfected siRNAs, shRNAs are endogenously transcribed. The eukaryotic H1 and U6 pol III promoters are the most commonly used to transcribe shRNAs. These promoters generate high levels of small, non-coding RNA transcripts and show high activity in all cell types.

For all these reasons it was decided to use the pSilencer 2.1-U6 puro siRNA Expression Vector (Ambion) to knockdown BGL3 expression via the RNAi pathway within the K562 cells (see section 2.3.16). This vector expresses siRNAs within mammalian cells using a U6 pol III promoter and permits selection and long-term RNAi experiments through an antibiotic resistance gene. To elicit BGL3 silencing, short hairpin RNAs (shRNAs) targeting the non-coding transcript were cloned into the vector downstream of the promoter. Once transfected into the K562 cell line, the insert-containing vector expresses the shRNA, which is rapidly processed by the cellular machinery into siRNA.

3.4.1 Hairpin siRNA Template Oligonucleotide Design

The first step to knockdown the BGL3 transcript using the pSilencer 2.1-U6 puro siRNA Expression Vector was to design complementary 60-64 mer oligonucleotides for the BGL3 sense and antisense transcripts (see sections 2.3.18 and 2.3.19). These oligonucleotides were later ligated into the pSilencer 2.1-U6 puro siRNA Expression Vector. The oligonucleotides encode 21-mer siRNAs specific to the BGL3 targets, a loop sequence separating the two complementary domains, and a polythymidine tract to terminate transcription and 3' overhanging UU dinucleotides to enhance their effectiveness. Addition of GGAA immediately downstream of the RNA pol III terminator site improves the efficiency of gene silencing, though the mechanism for this effect is unknown[103]. All designed siRNAs have *Bam* HI and *Hind* III to facilitate directional cloning into the pSilencer 2.1-U6 puro siRNA Expression Vector (Figure 3-12).

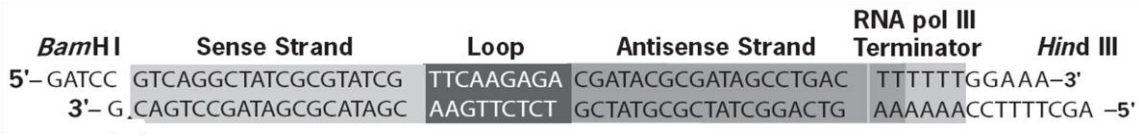


Figure 3- 12 Schematic overview of the designed oligonucleotides to be cloned into the pSilencer 2.1-U6 puro siRNA Expression Vector.

Adapted from: Ambion, 2008, p. 9[103].

Six different siRNAs were designed to target different regions of the BGL3 sense and antisense transcripts, in order to increase the probabilities to effectively target the non-coding transcript. Furthermore, siRNAs were designed to target different regions along the length of the BGL3 sense and antisense transcripts, to increase the chances of targeting a region of BGL3 without significant higher order structures or regulatory proteins bound (Figures 3-4 and 3-5). siRNAs were also selected with G/C contents ranging from 30 to 50%, which are more active than those with a higher G/C content.

A) Sense BGL3 siRNAs

Four shRNAs were designed for the sense transcript of BGL3. Figure 3-13 displays the four siRNA targets selected using Ambion's Target Finder.

>Sense BGL3 transcript

TTAGCAGTAACTGCTGAATTCCTGGTTGGCTGATGGAAAGATGGGGCAGCTGTTCACTGGTACGCAGGGT
TTTAGATGTATGTACCT**AAGGATATGAGGTATGGCAAT**GAACAGAAATTCTTTTGGGAATGAGTTTTAGG
GCCATTAAAGGACATGACCTGAAGTTTCCTCTGAGGCCAGTCCCCACAACTCAATATAAATGTGTTTCCT
GCATATAGTCAAAGTTGCCACTTCTTTTTCTTCATATCATCGATCTCTGCTCTTAAAGATAATCTTGGTT
TTGCCTC**AAACTGTTTGTCACTACAAAC**TTTCCCCATGTTCCTAAGTAAAACAGGTAACTGCCTCTCAAC
TATATCAAGTAGACTAAAATATTGTGTCTCTAATATCAGAAATTCAGCTTTAATATATTGGGTTTAACTC
TTTGAAATTTAGAGTCTCCTTGAAATACACATGGGGGTGATTTCCTAAACTTTATTCTTGT**AAGGATTTA**
TCTCAGGGGTAACACACAAACCAGCATCCTGAACCTCTAAGTATGAGGACAGTAAGCCTTAAGAATATAA
AATAAACTGTTCTTCTCTCTGCCGGTGGAAGTGTGCCCTGTCTATTCCTGAAATTGCTTGTTTGAGACGC
ATGAGACGTGCAGCACATGAGACACGTGCAGCAGCCTGTGGAATATTGTCAGTGAAGAATGTCTTTGCCT
GATTAGATATAAAGACAAGTTAAACACAGCATTAGACTATAGATCAAGCCTGTGCCAGACAC**AAATGACC**
TAATGCCCAGCACGGGCCACGGAATCTCCTATCCTCTTGCTTGAACAGAGCAGCACACTTCTCCCCCAAC
ACTATTAGATGTTCTGGCATAATTTTGTAGATATGTAGGATTTGACATGGACTATTGTTCAATGATTCAG
AGGAAATCTCCTTTGTTCAGATAAGTACACTGACTACTAAATGGATTAAAAAACACAGTAATAAAACCCA
GTTTTCCCCTTAAAAAAAAAAAAAAAAAAAAAAAAA

Figure 3- 13 Selected siRNA target sites (in red) for the sense BGL3 transcript.

shRNAs were designed using Ambion's Design tool for the pSilencer Vectors. The details of the design of the shRNAs for these siRNA target sites are indicated below. siRNA target sequence:

1. AAGGATATGAGGTATGGCAAT

   Position in gene sequence: 88

   GC content: 38.1%

   Sense strand siRNA: GGAUAUGAGGUAUGGCAAUUU

   Antisense strand siRNA: AUUGCCAUACCUCAUAUCCUU

   shRNA design:

```
   BamH1    Sense Strand       Loop      Antisense strand              Hind III
5'-GATCC GGATATGAGGTATGGCAAT TTCAAGAGA ATTGCCATACCTCATATCC TTTTTT GGAA A-3'
    3'-G CCTATACTCCATACCGTTA AAGTTCTCT TAACGGTATGGAGTATAGG AAAAAA CCTT TTCGA-5'
```

2. AAACTGTTTGTCACTACAAAC

   Position in gene sequence: 288

   GC content: 33.3%

   Sense strand siRNA: ACUGUUUGUCACUACAAACUU

   Antisense strand siRNA: GUUUGUAGUGACAAACAGUUU

   shRNA design:

```
   BamH1    Sense Strand Loop     Antisense strand               Hind III
5'-GATCC ACTGTTTGTCACTACAAAC TTCAAGAGA GTTTGTAGTGACAAACAG TTTTTTT GGAA A-3'
    3'-G TGACAAACAGTGATGTTTG AAGTTCTCT CAAACATCACTGTTTGTC AAAAAAA CCTT TTCGA G-5'
```

3. AAGGATTTATCTCAGGGGTAA

   Position in gene sequence: 482

   GC content: 38.1%

   Sense strand siRNA: GGAUUUAUCUCAGGGGUAAUU

   Antisense strand siRNA: UUACCCCUGAGAUAAAUCCUU

   shRNA design:

```
   BamH1    Sense Strand        Loop     Antisense strand               Hind III
5'-GATCC GGATTTATCTCAGGGGTAA TTCAAGAGA TTACCCCTGAGATAAATCC TTTTTT GGAA A-3'
    3'-G CCTAAATAGAGTCCCCATT AAGTTCTCT AATGGGGACTCTATTTAGG AAAAAA CCTT TTCGA -5'
```

4. AAATGACCTAATGCCCAGCAC

   Position in gene sequence: 763

   GC content: 47.6%

   Sense strand siRNA: AUGACCUAAUGCCCAGCACUU

   Antisense strand siRNA: GUGCUGGGCAUUAGGUCAUUU

   shRNA design:

```
   BamH1    Sense Strand        Loop     Antisense strand               Hind III
5'-GATCC ATGACCTAATGCCCAGCAC TTCAAGAGA GTGCTGGGCATTAGGTCA TTTTTTT GGAA A-3'
    3'-G TACTGGATTACGGGTCGTG AAGTTCTCT CACGACCCGTAATCCAGT AAAAAAA CCTT TTCGA-5'
```

All these siRNA target sequences are specific for the BGL3 transcript and do not have homology with any other human gene, as confirmed using the BLAT tool.

B) Antisense BGL3 siRNAs

Using the same procedure described above, two shRNAs were designed for the antisense transcript of BGL3. Figure 3-14 shows the two siRNA targets selected form the report of the Ambion's Target Finder.

>Antisense BGL3 transcript

TTTTTTTTTTTTTTTTTTTTTTTTTTTAAGGGGAAAACTGGGTTTTATTACTGTGTTTTTTAATCCATTT
AGTAGTCAGTGTACTTATCTGAACAAAGGAGATTTCCTCTGAATCATTGAACAATAGTCCATGTCAAATC
CTACATATCTACAAAATTATGCCAGAACATCTAATAGTGTTGGGGGAGAAGTGTGCTGCTCTGTTCAAGC
AAGAGGATAGGAGATTCCGTGGCCCGTGCTGGGCATTAGGTCATTTGTGTCTGGCACAGGCTTGATCTAT
AGTCTAATGCTGTGTTTAACTTGTCTTTATATCTAATCAGGCAAAGACATTCTTCACTGAC**AATATTCCA**
CAGGCTGCTGCACGTGTCTCATGTGCTGCACGTCTCATGCGTCTCAAACAAGCAATTTCAGGAATAGACA
GGGCACACTTCCACCGGCAGAGAGAAGAACAGTTTATTTTATATTCTTAAGGCTTACTGTCCTCATACTT
AGAGGTTCAGGATGCTGGTTTGTGTGTTACCCCTGAGATAAATCCTTACAAGAATAAAGTTTAGGAAATC
ACCCCCATGTGTATTTCAAGGAGACTCTAAATTTCAAAGAGTTAAACCCAATATATTAAAGCTGAATTTC
TGATATTAGAGACACAATATTTTAGTCTACTTGATATAGTTGAGAGGCAGTTACCTGTTTTACTTAGG**AA**
CATGGGGAAAGTTTGTAGTGACAAACAGTTTGAGGCAAAACCAAGATTATCTTTAAGAGCAGAGATCGAT
GATATGAAGAAAAAGAAGTGGCAACTTTGACTATATGCAGGAAACACATTTATATTGAGTTGTGGGGACT
GGCCTCAGAGGAAACTTCAGGTCATGTCCTTTAATGGCCCTAAAACTCATTCCCAAAAGAATTTCTGTTC
ATTGCCATACCTCATATCCTTAGGTACATACATCTAAAACCCTGCGTACCAGTGAACAGCTGCCCCATCT
TTCCATCAGCCAACCAGGAATTCAGCAGTTACTGCTAA

Figure 3- 14 Selected siRNA target sites (in red) for the antisense BGL3 transcript.

Below appear the details of the designed shRNAs for the selected siRNA target sites. As before shRNAs were designed using the Ambion's Design tool for the pSilencer Vectors. siRNA target sequence:

1. AATATTCCACAGGCTGCTGCA

   Position in gene sequence: 342

   GC content: 47.6%

   Sense strand siRNA: UAUUCCACAGGCUGCUGCAUU

   Antisense strand siRNA: UGCAGCAGCCUGUGGAAUAUU

   shRNA design:

```
   BamH1    Sense Strand        Loop      Antisense strand              Hind III
5'-GATCC GTATTCCACAGGCTGCTGCA TTCAAGAGA TGCAGCAGCCTGTGGAATATT TTTT GGAA A-3'
    3'-G CATAAGGTGTCCGACGACGT AAGTTCTCT ACGTCGTCGGACACCTTATAA AAAA CCTT TTCGA-5'
```

2. AACATGGGGAAAGTTTGTAGT

   Position in gene sequence: 699

   GC content: 38.1%

   Sense strand siRNA: CAUGGGGAAAGUUUGUAGUUU

   Antisense strand siRNA: ACUACAAACUUUCCCCAUGUU

   shRNA design:

```
   BamH1    Sense Strand       Loop      Antisense strand                  Hind III
5'-GATCC GCATGGGGAAAGTTTGTAG TTTCAAGAG AACTACAAACTTTCCCCA TGTTTTT TGGAA A-3'
    3'-G CGTACCCCTTTCAAACATC AAAGTTCTC TTGATGTTTGAAAGGGGT ACAAAAA ACCTT TTCGA
```

BLAT analysis confirmed that all these siRNA target sequences are specific for the BGL3 transcript and do not have homology with any other region of the human genome.

C) BGL3 siRNAs UCSC custom track

A UCSC custom track was created to display the designed BGL3 shRNAs and show their targeted position within the BGL3 transcript (Figure 3-15).

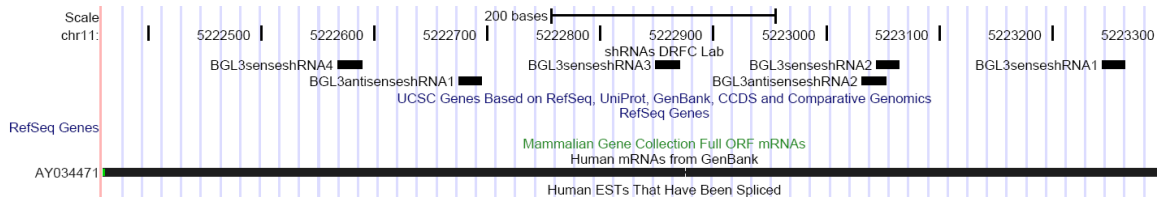


Figure 3- 15 UCSC custom track showing the designed siRNAs for the BGL3 transcript (AY034471).

3.4.2 Cloning strategy for the pSilencer 2.1-U6 puro siRNA Expression Vector

The cloning strategy of the designed shRNA inserts into the pSilencer 2.1-U6 puro siRNA Expression Vector was designed as follows (Figure 3-16). First, the plasmid was propagated and extracted from *E. coli*. Second, the scrambled shRNA insert was removed by digesting the plasmid with *Bam* HI and *Hind* III. Third, the linearised plasmid was purified from an agarose gel. Fourth, the designed shRNAs with compatible ends were annealed and ligated into the plasmid, and the constructs were

selected and propagated in *E. coli*. Finally, extracted plasmids were used to transfect K562 cells.

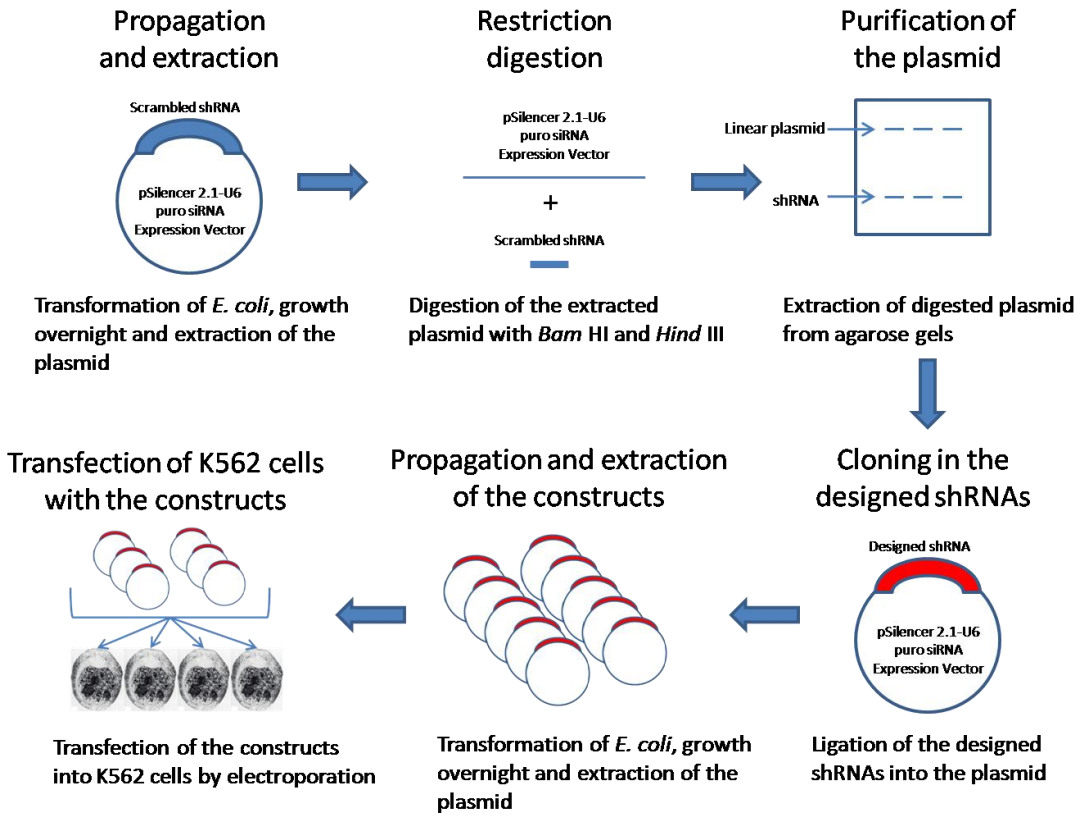


Figure 3- 16 Schematic overview of the cloning strategy for the pSilencer 2.1-U6 puro plasmid.

3.4.3 Propagation and extraction of pSilencer 2.1-U6 puro plasmid

pSilencer 2.1-U6 puro siRNA Expression Vector was kindly provided by Peter Fraser (Babraham Institute, Cambridge). The plasmid was in a circular configuration and contained a scrambled shRNA insert without significant homology to any gene of the human genome. Chemically competent K12 *E.coli* cells were transformed with 5 µL of the circular pSilencer 2.1-U6 puro plasmid to create stock of the plasmid (see section 2.3.8). Plasmid was then extracted from 5 mL overnight broth cultures with ampicillin using the QIAprep Spin Miniprep Kit (Qiagen) (see section 2.3.9). The extracted plasmid was then assessed on an agarose gel (see section 2.3.11). Electrophoresis produced one band of approximately 3500 bp in size (Figure 3-17).

3.4.4 Restriction digestion of the extracted pSilencer 2.1-U6 puro plasmid

The pSilencer 2.1-U6 puro plasmid was linearised with the restriction enzymes *Bam*H I and *Hind* III to remove the scrambled shRNA insert (see section 2.3.10). The plasmid was digested with the restriction enzymes *Bam*H I and *Hind* III. The resulting digests were electrophoresed on an agarose gel (Figure 3-17). Digestion of the plasmids produced one clear band of approximately 4500 bp in size and confirmed the presence of the pSilencer 2.1-U6 puro plasmid which is 4455 bp in length.

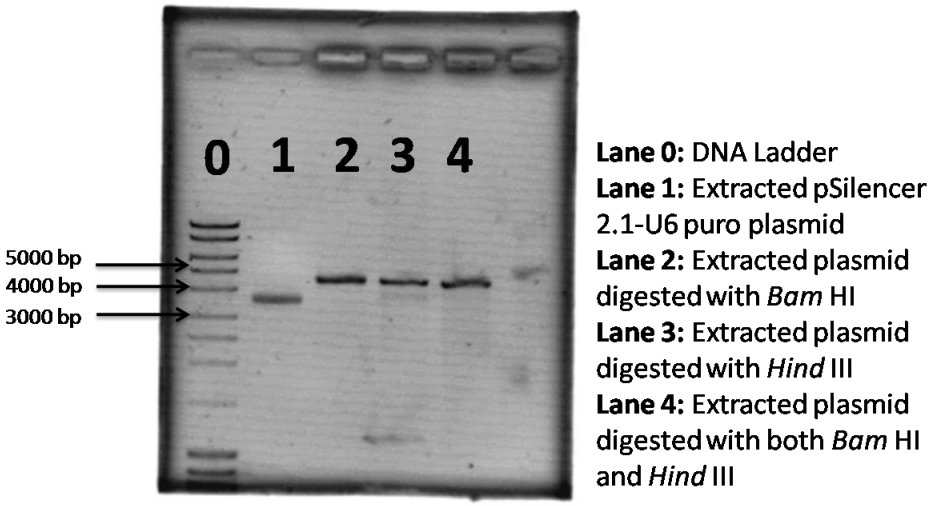


Figure 3- 17 Gel electrophoresis of digested pSilencer 2.1-U6 puro plasmid.

7 µL of extracted and digested plasmids were run on a 1.2% agarose gel for 45 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). The extracted plasmid produced one clear band of approximately 3500 bp in size. The digested plasmids produced one clear band of approximately 4500 bp in size.

3.4.5 Extraction of digested pSilencer 2.1-U6 puro plasmid from agarose gels

In order to generate sufficient linear plasmid for subsequent ligation reactions, 6 additional digestions of the extracted plasmid were performed using both *Bam*H I and *Hind* III restriction enzymes. The resulting digests were run in an agarose gel (Figure 3-18).

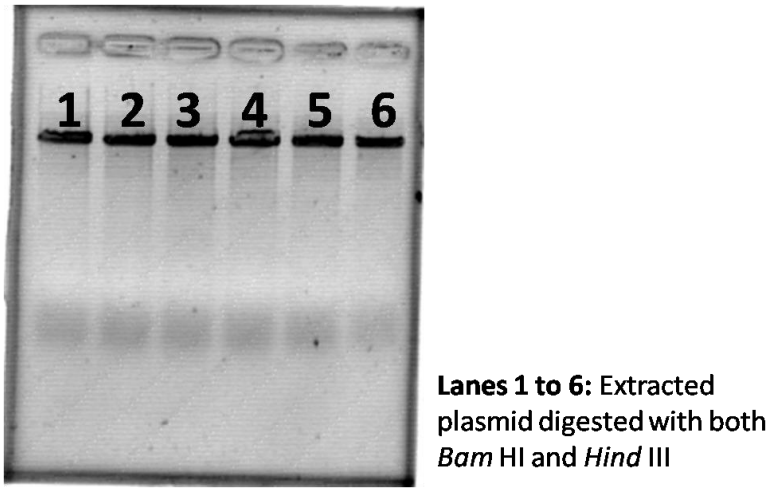


Figure 3- 18 Gel electrophoresis of digested pSilencer 2.1-U6 puro plasmid.

30 µL of extracted and digested plasmids were run on a 1.2% agarose gel for 30 minutes at 100 V. The digested plasmids produced one clear band.

The electrophoresed digests were subsequently excised and extracted from the agarose gel using the QIAquick Gel Extraction Kit (see section 2.3.12). The purified linear pSilencer 2.1-U6 puro plasmid was subsequently run on an agarose gel to assess the recovery of the linear plasmid (Figure 3-19).

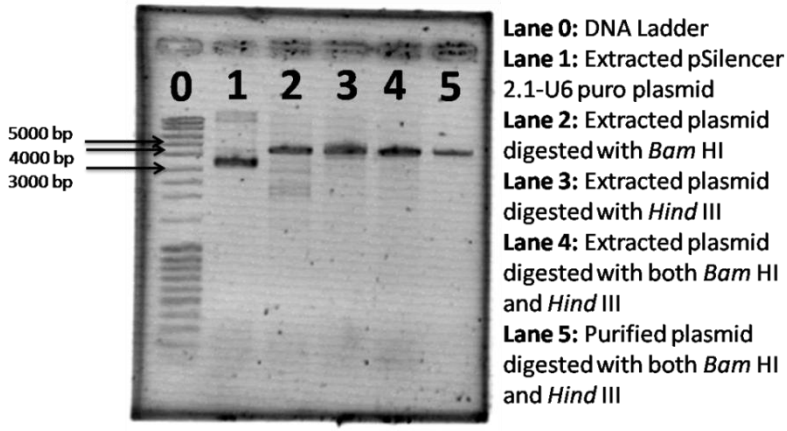


Figure 3- 19 Gel electrophoresis of digested and purified pSilencer 2.1-U6 puro plasmid.

7 µL of extracted, digested and purified plasmids were run on a 1.2% agarose gel for 30 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). The extracted plasmid produced one clear band of approximately 3500 bp in size. The digested and purified plasmids produced one clear band of approximately 4500 bp in size.

3.4.6 Ligation of the shRNA oligonucleotides into the pSilencer 2.1-U6 puro plasmid

Prior to the ligation reactions, the approximate DNA concentration of the purified linear pSilencer 2.1-U6 puro plasmid was estimated to be 25 ng/µL using a spectrophotometer. Since the synthesised shRNA inserts were diluted to 8 ng/µL, the volumes of the individual inserts and plasmid required were worked out for optimal reaction ligations at a 3:1 molar ratio of insert to plasmid.

A) Calculation of molar ratios of purified linear pSilencer 2.1-U6 puro plasmid to shRNA insert

The initial molar ratio of plasmid to insert was:

$$\frac{Plasmid\ bp}{Insert\ bp} = \frac{4455\ bp}{64\ bp} = 69.6$$

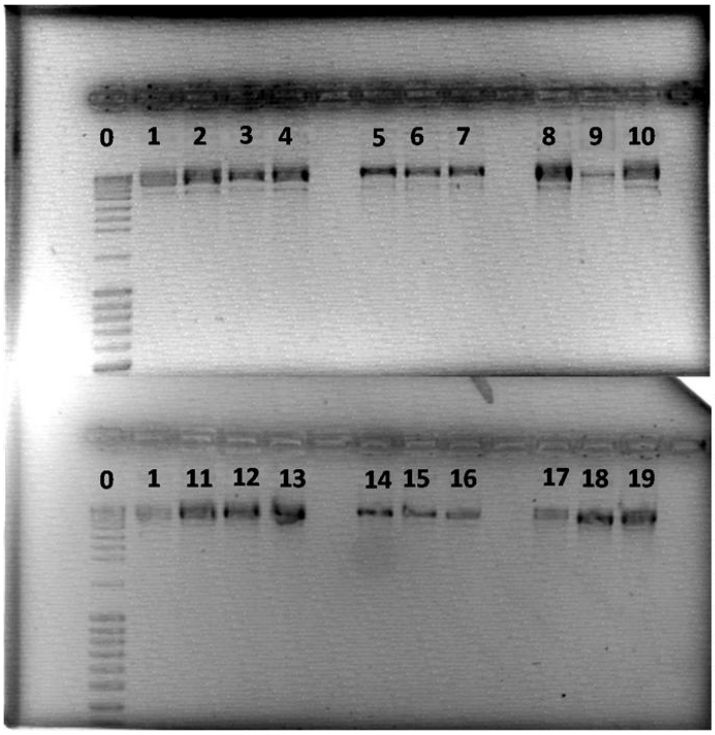
Dividing 69.6 by 3 it provides a 3:1 shRNA insert to purified linear pSilencer 2.1-U6 puro plasmid of 23.2. Therefore 1 ng of shRNA insert to 23.2 ng of plasmid is equivalent to a 3:1 molar ratio of insert to plasmid. Finally, referring to the concentrations of plasmid and insert indicated above, 1 µL of the diluted shRNA insert requires approximately 7 µL of purified linear pSilencer 2.1-U6 puro plasmid to give a 3:1 molar ratio.

The final volume of the ligation reactions was adjusted to 10 µL adding 1 µL T4 DNA ligase and 1 µL T4 DNA ligase buffer (see section 2.3.21). This ligation procedure was repeated for each of the 6 designed shRNAs to target the BGL3 transcript.

3.4.7 Transformation of chemically competent K12 E. coli cells with ligated pSilencer 2.1-U6 puro plasmid

Chemically competent K12 *E.coli* cells were transformed with 10 µL of ligated pSilencer 2.1-U6 puro plasmid (see section 2.3.8). This process was repeated for the six ligated plasmids. For each transformation, three colonies were picked-up and growth on LB broth in the presence of ampicillin. No colonies were detected in the minus-insert ligation control cultures. Plasmids were then extracted from the resulting 18 overnight 5 mL broth cultures with ampicillin using the QIAprep Spin Miniprep Kit (see section

2.3.9). The extracted plasmids were then assessed on an agarose gel (see section 2.3.11) (Figure 3-20).



Lane	Content
0	DNA Ladder
1	Scrambled insert
2	Sense BGL3 shRNA 1 clone A
3	Sense BGL3 shRNA 1 clone B
4	Sense BGL3 shRNA 1 clone C
5	Sense BGL3 shRNA 2 clone A
6	Sense BGL3 shRNA 2 clone B
7	Sense BGL3 shRNA 2 clone C
8	Sense BGL3 shRNA 3 clone A
9	Sense BGL3 shRNA 3 clone B
10	Sense BGL3 shRNA 3 clone C
11	Sense BGL3 shRNA 4 clone A
12	Sense BGL3 shRNA 4 clone B
13	Sense BGL3 shRNA 4 clone C
14	Antisense BGL3 shRNA 1 clone A
15	Antisense BGL3 shRNA 1 clone B
16	Antisense BGL3 shRNA 1 clone C
17	Antisense BGL3 shRNA 2 clone A
18	Antisense BGL3 shRNA 2 clone B
19	Antisense BGL3 shRNA 2 clone C

Figure 3- 20 Gel electrophoresis of ligated pSilencer 2.1-U6 puro plasmids.

7 µL of extracted plasmids were run on a 1.2% agarose gel for 30 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). The extracted plasmids produced one clear band of approximately 3500 bp in size.

All the plasmid extractions yielded strong signal in the agarose gels. Bacterial clones and the extracted plasmids were kept frozen as glycerol stocks for further analysis.

3.4.8 Verification of the presence of the shRNA insert in the extracted pSilencer 2.1-U6 puro plasmid

Extracted pSilencer 2.1-U6 puro plasmids were sent for sequencing to check if the transformed *E. coli* actually contained the plasmid with the insert correctly ligated and that there were no unwanted mutations (see section 2.3.22). 10 µL of extracted plasmid of one randomly selected clone of each shRNA insert was sequenced. Sequencing reactions confirmed the presence of the correct shRNA inserts in the clone A of the sense BGL3 shRNA 3, clone C of the sense BGL3 shRNA 4 and clone A of the

antisense BGL3 shRNA 1. Sequencing reactions for the clone A of the sense BGL3 shRNA 1, clone C of the sense BGL3 shRNA 2 and clone C of the antisense BGL3 shRNA 2 generated short reads that prevented the confirmation of the presence of the shRNA inserts. This could be caused by the secondary structure of the shRNAs, which may be responsible for preventing the extension of the read.

Furthermore, the original pSilencer 2.1-U6 puro plasmid was also sent for sequencing in order to determine the sequence of the scrambled insert that will be used as a control. BLAT analysis of the sense strand of the siRNA contained in the shRNA scrambled sequence (5'-GATCCACTACCGTTGTTATAGGTGTTCAAGAGACACCTATAAC AACGGTAGTTTTTTGGAAA-3') showed that the scrambled shRNA insert does not have homology with any region of the human genome.

3.5 Generation of pEF6/V5-His A plasmid constructs

In this project it was also attempted to alter the levels of the BGL3 transcript by cloning it into the constitutively expressed pEF6/V5-His A plasmid (Invitrogen) in order to observe if the over-expression of the non-coding transcript perturbs the expression levels of the β- and γ-globin genes within the human β-globin locus in K562 cells. Although the pEF6/V5-His A plasmid is designed for overproduction of recombinant proteins, this plasmid was selected to over-express the BGL3 transcript because it contains the human elongation factor 1α-subunit promoter (hEF-1α). This promoter provides high-level expression across a broad range of species and cell types. The pEF6/V5-His A plasmid also contains the blasticidin resistance gene (*bsd*), which allows selection of stable cell lines using blasticidin in the culture medium.

3.5.1 Primers design

The first step to clone the BGL3 transcript into the pEF6/V5-His A plasmid was to amplify the cDNA of the non-coding transcript. Paired primers were designed to amplify the BGL3 transcript. In the Figure 3-21 it is indicated the region of the BGL3 cDNA amplified by the primers.

> Sense BGL3 transcript

AAGGGGAAAACTGGGTTTTATTACTGTGTTTTTTAATCCATTTAGTAGTCAGTGTACTTATCTGAACAAA
GGAGATTTCCTCTGAATCATTGAACAATAGTCCATGTCAAATCCTACATATCTACAAAATTATGCCAGAA
CATCTAATAGTGTTGGGGGAGAAGTGTGCTGCTCTGTTCAAGCAAGAGGATAGGAGATTCCGTGGCCCGT
GCTGGGCATTAGGTCATTTGTGTCTGGCACAGGCTTGATCTATAGTCTAATGCTGTGTTTAACTTGTCTT
TATATCTAATCAGGCAAAGACATTCTTCACTGACAATATTCCACAGGCTGCTGCACGTGTCTCATGTGCT
GCACGTCTCATGCGTCTCAAACAAGCAATTTCAGGAATAGACAGGGCACACTTCCACCGGCAGAGAGAAG
AACAGTTTATTTTATATTCTTAAGGCTTACTGTCCTCATACTTAGAGGTTCAGGATGCTGGTTTGTGTGT
TACCCCTGAGATAAATCCTTACAAGAAATAAAGTTTAGGAAATCACCCCCATGTGTATTTCAAGGAGACT
CTAAATTTCAAAGAGTTAAACCCAATATATTAAAGCTGAATTTCTGATATTAGAGACACAATATTTTAGT
CTACTTGATATAGTTGAGAGGCAGTTACCTGTTTTACTTAGGAACATGGGGAAAGTTTGTAGTGACAAAC
AGTTTGAGGCAAAACCAAGATTATCTTTAAGAGCAGAGATCGATGATATGAAGAAAAAGAAGTGGCAACT
TTGACTATATGCAGGAAACACATTTATATTGAGTTGTGGGGACTGGCCTCAGAGGAAACTTCAGGTCATG
TCCTTTAATGGCCCTAAAACTCATTCCCAAAAGAATTTCTGTTCATTGCCATACCTCATATCCTTAGGTA
CATACATCTAAAACCCTGCGTACCAGTGAACAGCTGCCCCATCTTTCCATCAGCCAACCAGGAATTCAGC
AGTTACTGCTAA

Forward primer: 5′-GGGAAAACTGGGTTTTATTAC-3′
Reverse primer: 5′-TTAGCAGTAACTGCTGAATTC-3′

Figure 3- 21 Region of the BGL3 cDNA amplified by the designed primers.

3.5.2 Addition of restriction enzyme sites to the primers

Bam HI and *Xba* I recognition sites were added onto the 5’ end of the primers to generate PCR products with cohesive ends for ligation into the pEF6/V5-His A plasmid. Furthermore, working with two different recognition sites gave us the possibility to ligate the insert into the plasmid in both sense and antisense directions (Figure 3-22).

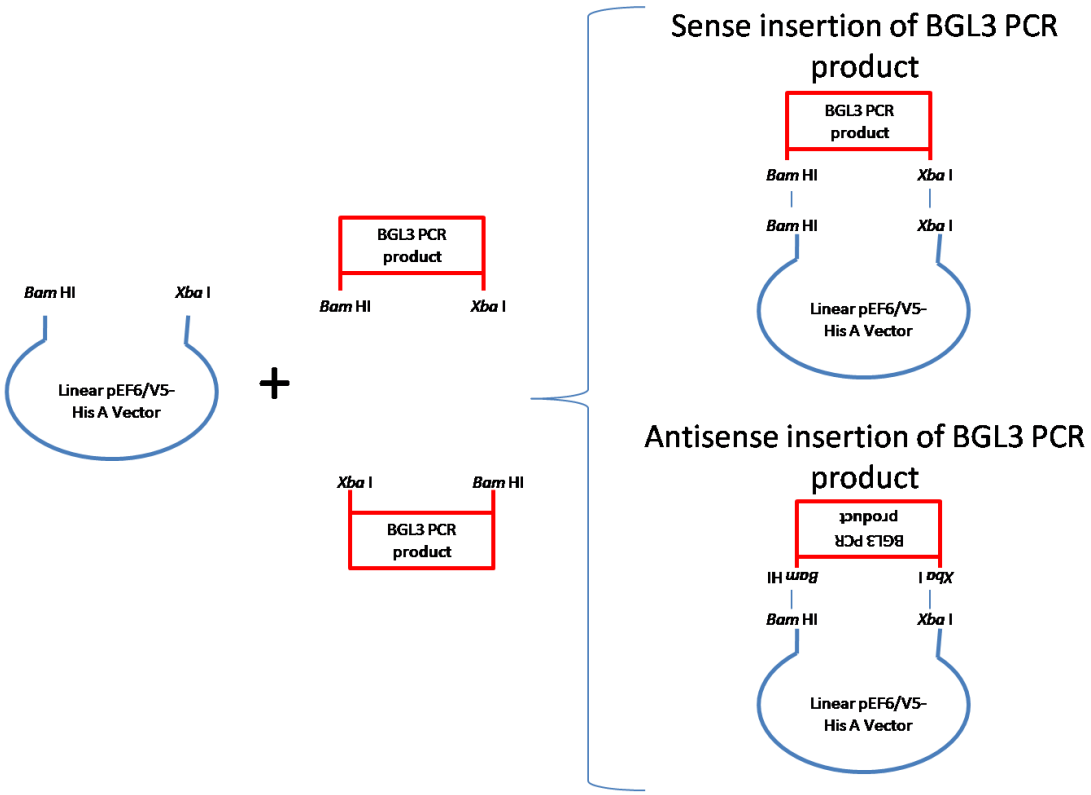


Figure 3- 22 Schematic overview of the strategy for the sense and antisense insertion of the BGL3 PCR products into the pEF6/V5-His A plasmid.

Since the cleavage sites were added at the end of the PCR products, the recognition sites were altered by adding additional bases in order to allow efficient cleavage. In the case of the *Bam* HI recognition site (5'-GGATCC-3'), two bases were added at each end resulting in the recognition site 5'-CGGGATCCCG-3'. The same procedure was followed in the case of the recognition site of *Xba* I (5'-TCTAGA-3'), in which two bases were added at each end resulting in the recognition site 5'-GCTCTAGAGC-3'.

The two recognition sites indicated above were added to the 5' end of the designed primer as shown in the table below.

Primer name	Direction	Sequence (5'->3')
Sense BGL3 insertion For	Forward	CGGGATCCCGGGGAAAACTGGGTT TTATTAC
Sense BGL3 insertion Rev	Reverse	GCTCTAGAGCTTAGCAGTAACTGCT GAATTC
Antisense BGL3 insertion For	Forward	GCTCTAGAGCGGGAAAACTGGGTT TTATTAC
Antisense BGL3 insertion For	Reverse	CGGGATCCCGTTAGCAGTAACTGC TGAATTC

The expected PCR products were 1012 bp in length for each primer pair. The NEBcutter tool confirmed that the BGL3 sequence does not contain any recognition sites for the Bam HI and Xba I restriction enzymes (Figure 3-23).

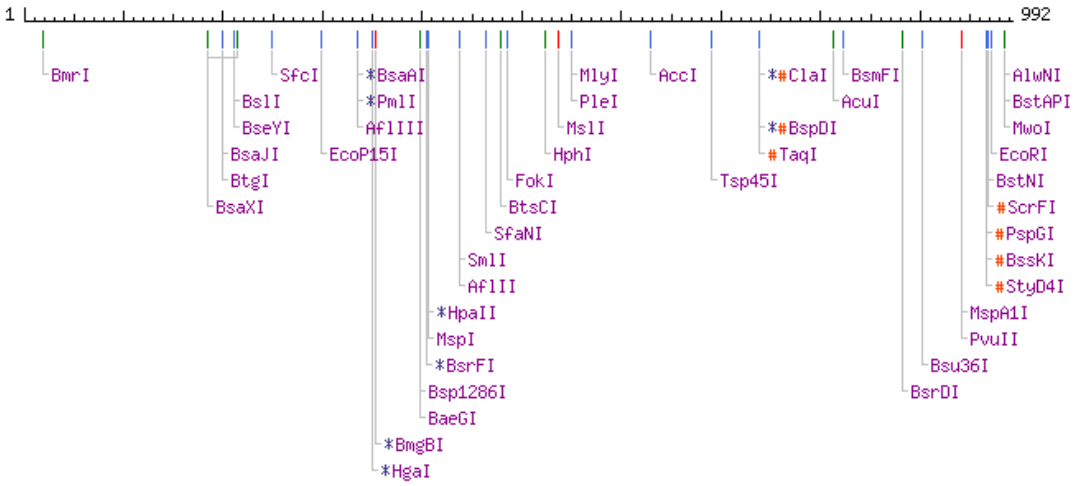


Figure 3- 23 Map of the restriction enzyme recognition sites present in the BGL3 DNA sequence.

3.5.3 Amplification of the BGL3 transcript

Using the designed primers indicated above, the BGL3 cDNA was amplified by PCR (see section 2.3.6). Negative controls, including a non-template controls and the –RT controls, were run for each reaction Gel electrophoresis (see section 2.3.11) showed PCR products of approximately 1000 bp in length, a size in accordance to the expected 1012 bp products (Figure 3-24; Lanes 1 and 5).

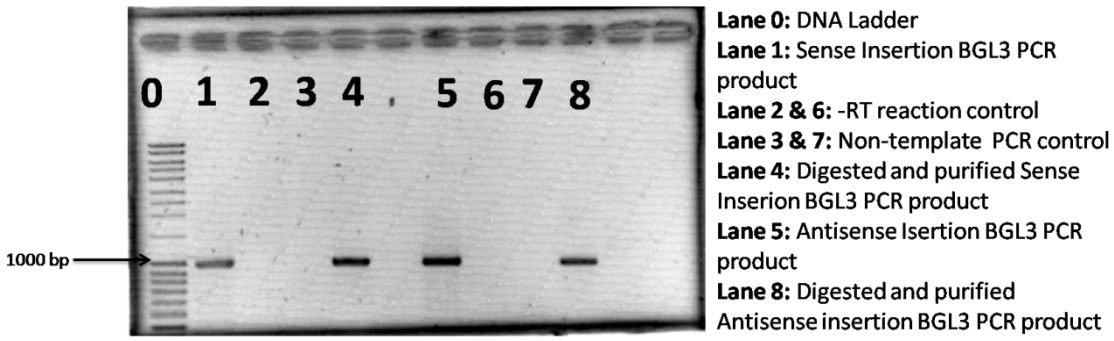


Figure 3- 24 Gel electrophoresis of PCR products of BGL3 transcripts for sense and antisense insertion into the pEF6/V5-His A plasmid.

7 µL of PCR products and digested and purified products were run on a 1.2% agarose gel for 45 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). Negative controls confirmed the absence of contamination or residual genomic DNA. PCR reactions produced one clear band of approximately 1000 bp in size. The digested and purified PCR products also produced one clear band of approximately 1000 bp in size.

3.5.4 Restriction digestion of PCR products

Once it was confirmed the correct amplification of the BGL3 transcripts by PCR, four additional PCR reactions were performed for both BGL3 sense and antisense transcript insertions in order to generate sufficient stock for further ligation reactions. Following PCR, products were digested with *Bam* HI and *Xba* I (see section 2.3.10) in order to generate cohesive ends compatible with those of the linear pEF6/V5-His A plasmid (see below). Digests were then electrophoresed in an agarose gel (see section 2.3.11) (Figure 3-25).

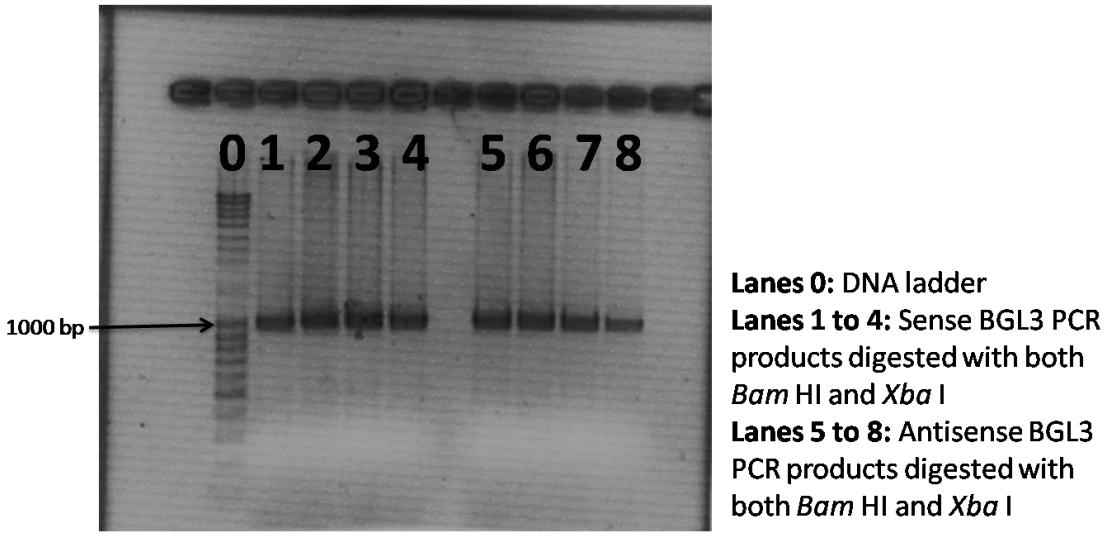


Figure 3- 25 Gel electrophoresis of the digested PCR products of BGL3 sense and antisense transcripts.

30 µL of digested PCR products were run on a 1.2% agarose gel for 45 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). The digested PCR products produced one clear band of approximately 1000 bp in size.

3.5.5 Extraction of digested PCR products from agarose gels

The electrophoresed digests were subsequently excised and extracted from the agarose gel using the QIAquick Gel Extraction Kit (see section 2.3.12). The purified digested PCR products were subsequently run on an agarose gel (see section 2.3.11) to assess the recovery of the amplified transcripts (Figure 3-24; Lanes 4 and 8) and kept frozen for further ligation into the pEF6/V5-His A plasmid.

3.5.6 Cloning strategy for the pEF6/V5-His A plasmid

The cloning strategy of the BGL3 PCR products into the pEF6/V5-His A plasmid was designed as follows (Figure 3-26). First, the plasmid was propagated and extracted from *E. coli*. Second, the plasmid was digested with *Bam* HI and *Xba* I. Third, the linearised plasmid was purified from an agarose gel. Fourth, the BGL3 PCR products with compatible ends were ligated into the plasmid and the constructs were propagated in *E. coli*. Finally, extracted plasmids will be used to transfect K562 cells.

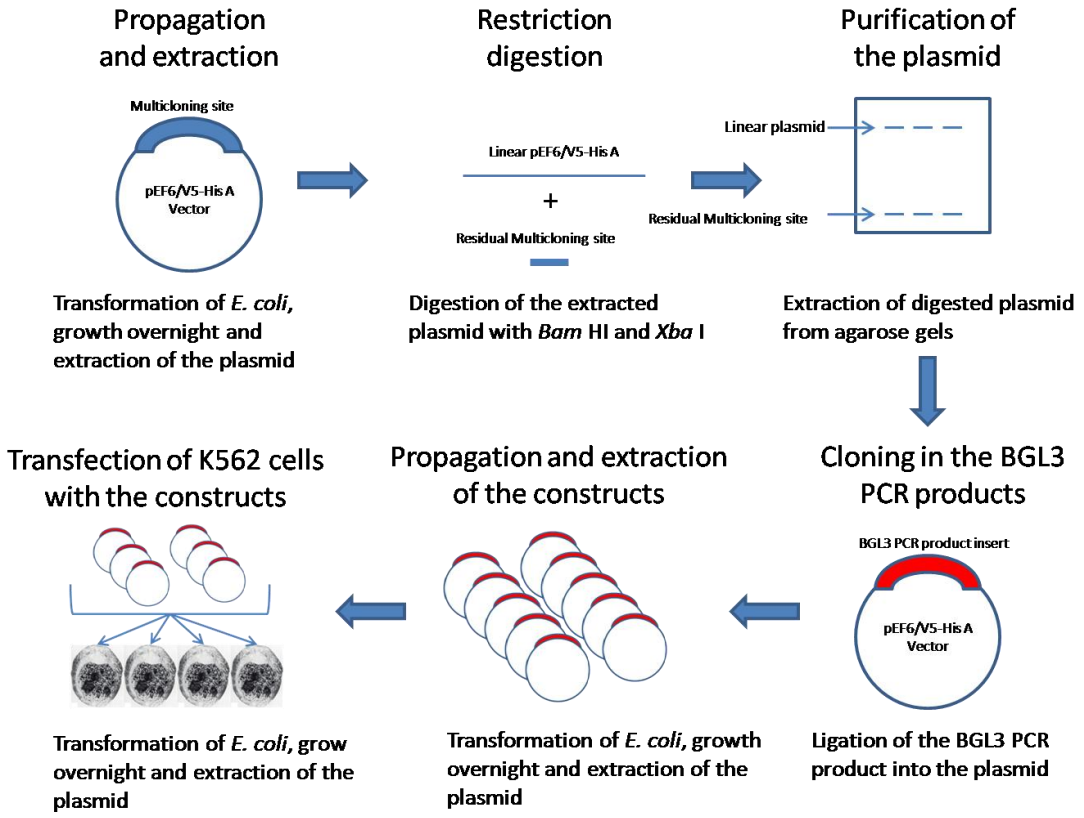


Figure 3- 26 Schematic overview of the cloning strategy for the pEF6/V5-His A plasmid.

3.5.7 Propagation and extraction of pEF6/V5-His A plasmid

pEF6/V5-His A plasmid was kindly provided by Stephen Best (Kings College, London). The plasmid was in a circular configuration and contained no insert. Chemically competent K12 *E.coli* cells were transformed with 5 µL of the circular pEF6/V5-His A plasmid to create stock of the plasmid (see section 2.3.8). Plasmid was then extracted from 5 mL overnight broth cultures with ampicillin using the QIAprep Spin Miniprep Kit (see section 2.3.9). The extracted plasmid was then assessed on an agarose gel (see section 2.3.11). Electrophoresis produced one band of approximately 5000 bp in size

3.5.8 Restriction digestion of the extracted pEF6/V5-His A plasmid

The pEF6/V5-His A plasmid was linearised to be used in further ligation reactions. The plasmid was digested with the restriction enzymes *Bam*H I and *Xba* I (see section 2.3.10). The resulting digests were electrophoresed on an agarose gel (see section

2.3.11) (Figure 3-27). Digestion of the plasmids produced one clear band of approximately 5800 bp in size and confirmed the presence of the pEF6/V5-His A plasmid which is 5800 bp in length.

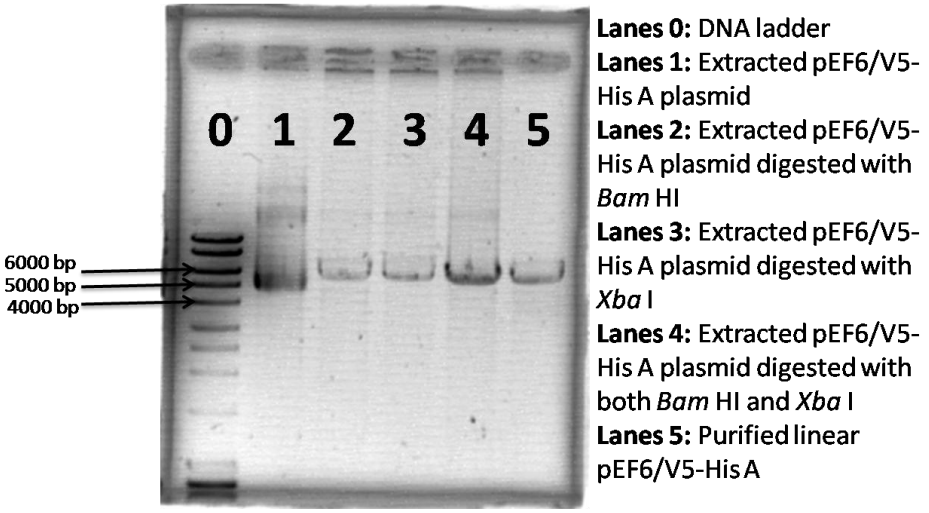


Figure 3- 27 Gel electrophoresis of extracted pEF6/V5-His A plasmid.

7 µL of extracted plasmid were run on a 1.2% agarose gel for 45 minutes with 5 µL of DNA ladder (range 80 – 10,000 bp). The extracted plasmid produced one clear band of approximately 5000 bp in size. The digested and purified plasmids produced bands of 5800 bp in size approximately.

3.5.9 Extraction of digested pEF6/V5-His A plasmid from agarose gels

In order to generate sufficient linear plasmid for further ligation reactions, 6 additional digestions of the extracted plasmid were performed using both *Bam*H I and *Xba* I restriction enzymes. The resulting digests were run on an agarose gel (see section 2.3.11) (Figure 3-28).

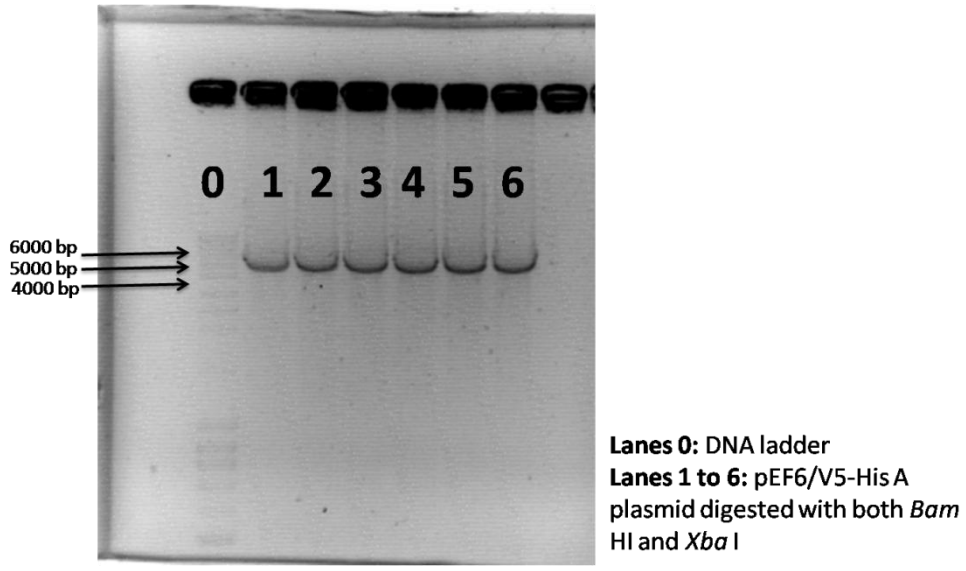


Figure 3- 28 Gel electrophoresis of digested pEF6/V5-His A plasmid.

30 µL of extracted and digested plasmids were run on a 1.2% agarose gel for 30 minutes. The digested plasmids produced one clear band of approximately 5800 bp in size.

The electrophoresed digests were subsequently excised and extracted from the agarose gel using the QIAquick Gel Extraction Kit (see section 2.3.12). The purified linear pEF6/V5-His A plasmid was subsequently run on an agarose gel to assess the recovery of the linear plasmid (see section 2.3.11) (Figure 3-27).

3.5.10 Ligation of the sense and antisense BGL3 transcripts into the pEF6/V5-His A plasmid

Prior to the ligation reactions, the approximate DNA concentration of the purified linear pEF6/V5-His A plasmid was estimated to be 15 ng/µL using a spectrophotometer. The concentration of the amplified BGL3 transcripts was also estimated using a spectrophotometer to be approximately 50 ng/µL. Thus, the volumes of the inserts and plasmid required were worked out for optimal reaction ligations at a 3:1 molar ratio of insert to plasmid.

A) Calculation of molar ratios of purified linear pEF6/V5-His A plasmid to BGL3 transcript insert

The initial molar ratio of plasmid to insert was:

$$\frac{Plasmid\ bp}{Insert\ bp} = \frac{5800\ bp}{1012\ bp} = 5.73$$

Dividing 5.73 by 3 it gives a 3:1 shRNA insert to purified linear pSilencer 2.1-U6 puro plasmid of 1.91. Therefore 1ng of insert to 1.91 ng of plasmid is equivalent to a 3:1 molar ratio of insert to plasmid. Finally, according to the concentrations of plasmid and insert indicated above, 1 µL of the diluted shRNA insert requires approximately 7 µL of purified linear pSilencer 2.1-U6 puro plasmid to give a 3:1 molar ratio.

The final volume of the ligation reactions was adjusted to 10 µL adding 1 µL T4 DNA ligase and 1 µL T4 DNA ligase buffer (see section 2.3.15). This ligation procedure was repeated for both sense and antisense insertions of the BGL3 transcript into the pEF6/V5-His A plasmid.

3.5.11 Transformation of chemically competent K12 E. coli cells with ligated pEF6/V5-His A plasmid

Chemically competent K12 *E.coli* cells were transformed with 10 µL of ligated pEF6/V5-His A plasmid (see section 2.3.8). This process was repeated for both plasmids containing the sense or antisense insertion of the BGL3 PCR product. For each transformation, three colonies were picked-up and growth on LB broth. Plasmids were then extracted from the resulting 6 overnight 5 mL broth cultures with ampicillin using the QIAprep Spin Miniprep Kit (see section 2.3.9). The extracted plasmids were then assessed on an agarose gel (see section 2.3.11) (Figures 3-30 and 3-31). All the plasmid extractions yielded strong signals in the agarose gels. Bacterial clones and extracted plasmids were stored frozen for further analysis.

3.5.12 Verification of the presence of the BGL3 inserts in the extracted pEF6/V5-His A plasmid

To check if the isolated colonies contained the ligated insert in the plasmid, plasmid extracts were digested with *Xba* I and *Kpn* I (see section 2.3.10). Recognition site for

Kpn I, as well as for the other restriction enzymes used, is within in the multicloning site of the pEF6/V5-His A plasmid (Figure 3-29).

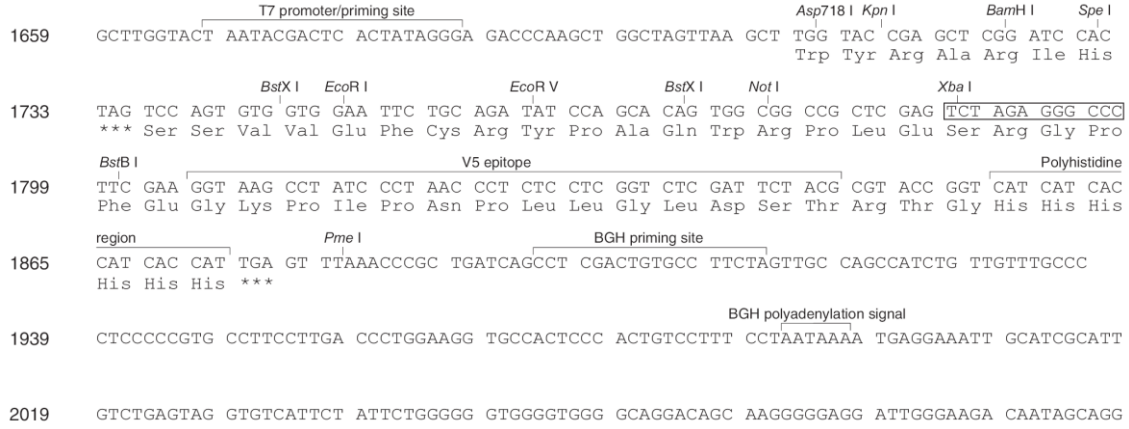
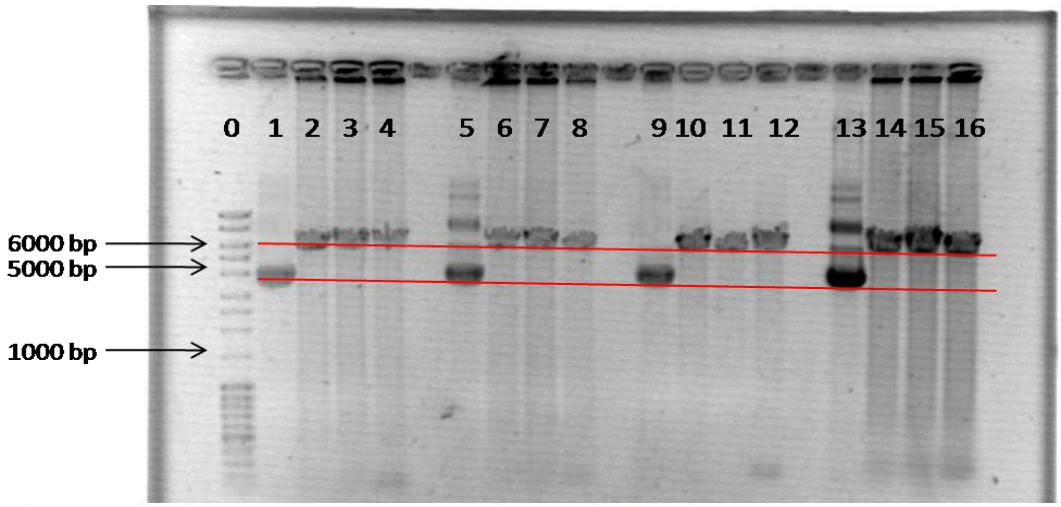


Figure 3- 29 Multicloning site of pEF6/V5-His A plasmid.

Restriction sites are labelled to indicate the cleavage site.



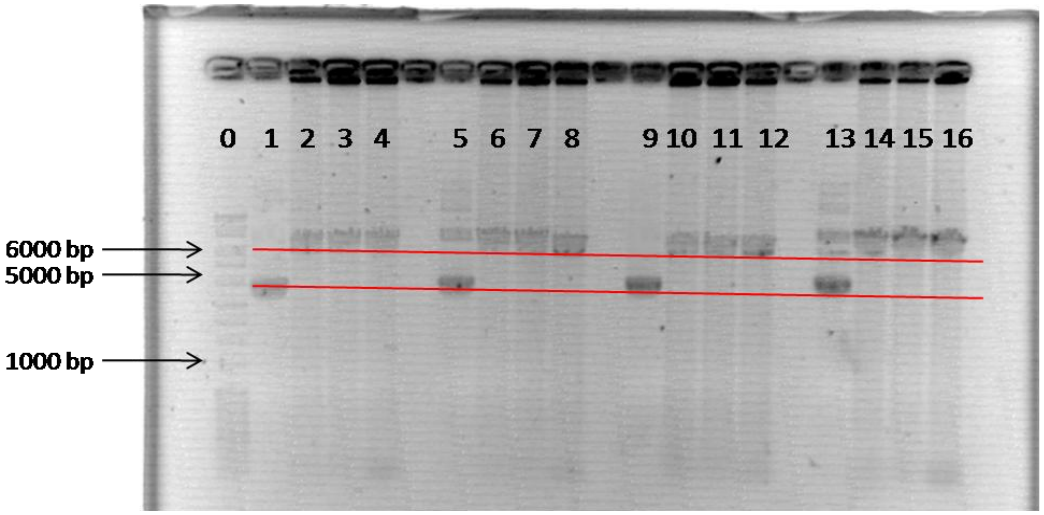
Lane	Content	Lane	Content
0	DNA ladder	9	pEF6/V5-His A plasmid with Sense BGL3 transcript clone B
1	pEF6/V5-His A plasmid	10	pEF6/V5-His A plasmid with Sense BGL3 transcript clone B digested with *Kpn* I
2	pEF6/V5-His A plasmid digested with *Kpn* I	11	pEF6/V5-His A plasmid with Sense BGL3 transcript clone B digested with *Xba* I
3	pEF6/V5-His A plasmid digested with *Xba* I	12	pEF6/V5-His A plasmid with Sense BGL3 transcript clone B digested with both *Kpn* I and *Xba* I
4	pEF6/V5-His A plasmid digested with both *Kpn* I and *Xba* I	13	pEF6/V5-His A plasmid with Sense BGL3 transcript clone C
5	pEF6/V5-His A plasmid with Sense BGL3 transcript clone A	14	pEF6/V5-His A plasmid with Sense BGL3 transcript clone C digested with *Kpn* I
6	pEF6/V5-His A plasmid with Sense BGL3 transcript clone A digested with *Kpn* I	15	pEF6/V5-His A plasmid with Sense BGL3 transcript clone C digested with *Xba* I
7	pEF6/V5-His A plasmid with Sense BGL3 transcript clone A digested with *Xba* I	16	pEF6/V5-His A plasmid with Sense BGL3 transcript clone C digested with both *Kpn* I and *Xba* I
8	pEF6/V5-His A plasmid with Sense BGL3 transcript clone A digested with both *Kpn* I and *Xba* I		

Figure 3- 30 Gel electrophoresis of ligated pEF6/V5-His A plasmids with the sense BGL3 insert.

7 µL of extracted, and extracted and digested plasmids were run on a 1.2% agarose gel for 100 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). All the

extracted plasmids produced one clear band of approximately 5000 bp in size. All the digested plasmids produced one clear band of approximately 5800 bp in size.

As shown in Figure 3-30, the cloning did not work in any of the three selected clones expected to contain the sense insertion of the BGL3 transcript. The digestion pattern of the ligated plasmid is identical to the digestion pattern of the control (neat pEF6/V5-His A plasmid; lanes 1 to 4). Furthermore, no bands are detected in the 1000 bp region in the lanes of the double digested plasmids. All this evidence indicates that the BGL3 sense insert (1000 bp in size approximately) was not ligated into the pEF6/V5-His A plasmid.



Lane	Content	Lane	Content
0	DNA ladder	9	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone B
1	pEF6/V5-His A plasmid	10	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone B digested with *Kpn* I
2	pEF6/V5-His A plasmid digested with *Kpn* I	11	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone B digested with *Xba* I
3	pEF6/V5-His A plasmid digested with *Xba* I	12	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone B digested with both *Kpn* I and *Xba* I
4	pEF6/V5-His A plasmid digested with both *Kpn* I and *Xba* I	13	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone C
5	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone A	14	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone C digested with *Kpn* I
6	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone A digested with *Kpn* I	15	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone C digested with *Xba* I
7	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone A digested with *Xba* I	16	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone C digested with both *Kpn* I and *Xba* I
8	pEF6/V5-His A plasmid with Antisense BGL3 transcript clone A digested with both *Kpn* I and *Xba* I		

Figure 3- 31 Gel electrophoresis of ligated pEF6/V5-His A plasmids with the antisense BGL3 insert. 7 µL of extracted, and extracted and digested plasmids were run on a 1.2% agarose gel for 100 minutes at 100 V with 5 µL of DNA ladder (range 80 – 10,000 bp). All the extracted plasmids produced one clear band of approximately 5000 bp in size. All the digested plasmids produced one clear band of approximately 5800 bp in size.

As shown in Figure 3-31 the digestion patterns of the ligated plasmids with the antisense insertion of the BGL3 PCR product are identical to the digestion pattern of the control (neat pEF6/V5-His A plasmid; lanes 1 to 4). No bands were detected in the 1000 bp region in the lanes of the double digested plasmids indicating that again, the cloning did not work in any of the three selected clones expected to contain the antisense BGL3 transcript insertion.

Due to time constrains, the cloning of the BGL3 PCR product was not pursued further. Attention was therefore focused on transfecting the confirmed pSilencer constructs into the K562 cells.

3.6 Antibiotic kill curves for K562 cells

In order to generate cell lines expressing the plasmid of interest, it is important to determine the minimum amount of antibiotic required to kill non-transfected cells. This was achieved generating antibiotic kill curves to determine the antibiotic dose-response of K562 cells. So it is necessary to identify the lowest level of antibiotic that kills non-transfected cells within approximately 3 days by testing antibiotic concentrations ranging from 0.01 to 10 µg/mL while keeping all other culture conditions equal.

3.6.1 Puromycin kill curve

The pSilencer 2.1-U6 puro vector contains the puromycin resistance gene *pac* from *Streptomyces alboniger*, which encodes puromycin N-acetyl-transferase and conveys resistance to puromycin to enable antibiotic selection in mammalian cells (see section 2.3.17). The puromycin dose-response of K562 cells is displayed in the Figure 3-32.

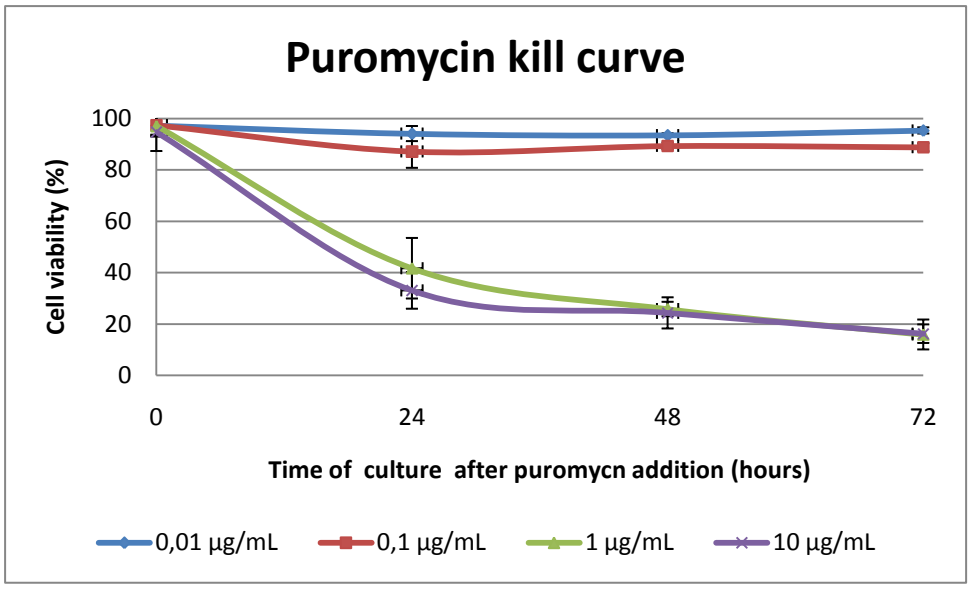


Figure 3- 32 Puromycin kill curve for K562 cells.

Cell viability was assessed using the Trypan Blue exclusion method. Data were obtained from two biological replicates for each time point after addition of puromycin to the K562 cell cultures. Error bars represent standard deviation.

As shown in Figure 3-32, the lowest puromycin concentration that gives massive cell death in 3 days is 1 µg/mL. This concentration was used to select K562 cells containing the pSilencer 2.1-U6 puro plasmid after transfection.

3.6.2 Blastidicin kill curve

The pEF6/V5-His A plasmid contains the blastidicin resistance gene *bsd* from *Streptomyces griseochromogenes*, which encodes a blastidicin S deaminase and conveys resistance to blastidicin to enable antibiotic selection in mammalian cells (see section 2.3.14). The blastidicin dose-response of K562 cells is displayed in the Figure 3-33.

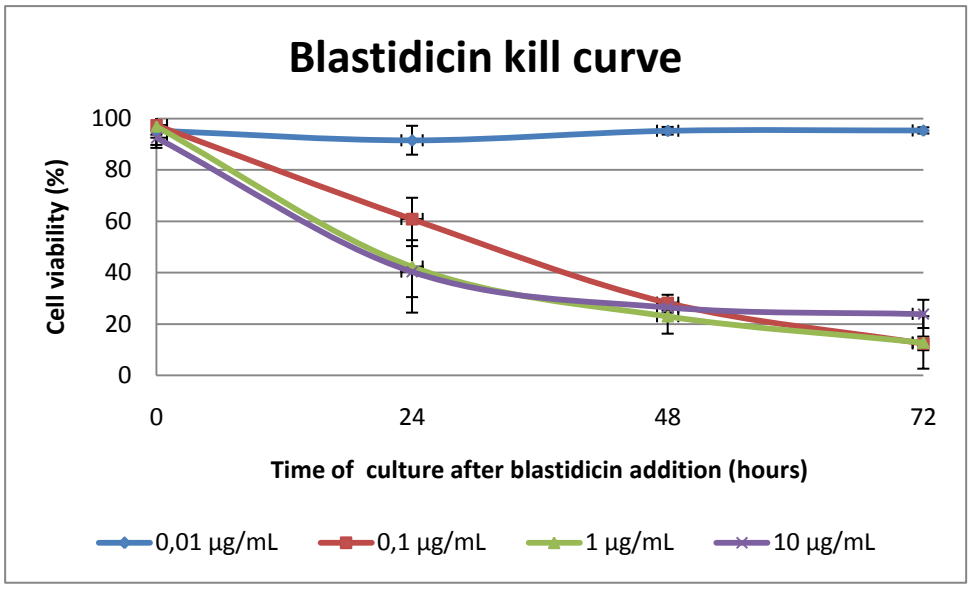


Figure 3- 33 Blastidicin kill curve for K562 cells.

Cell viability was assessed using the Trypan Blue exclusion method. Data were obtained from two biological replicates for each time point after addition of blastidicin to the K562 cell cultures. Error bars represent standard deviation.

As shown in Figure 3-33, the lowest blastidicin concentration that gives massive cell death in 3 days is 0.1 µg/mL. This concentration would have to be used to select K562 cells containing the pEF6/V5-His A plasmid after transfection.

3.7 Transient transfection of K562 cells with pSilencer constructs

The last step of this project involved transfecting the K562 cells with pSilencer constructs containing the inserts against the BGL3 transcript in order to actually perturb the expression levels of the non-coding transcript by knocking it down using the RNAi pathway (see section 2.3.23). The effect of the knockdown of the BGL3 transcript was tested on the activity of the γ- and β-globin genes, which were quantified using qRT-PCR, as well as the actual BGL3 transcript.

The pSilencer 2.1-U6 puro plasmids that were confirmed by sequencing to contain the correct shRNA insert were used to transiently transfect K562 cells. Thus, K562 cells were transfected with the pSilencer2.1-U6 puro plasmids containing the sense BGL3 shRNA 3, the sense BGL3 shRNA 4 and the antisense BGL3 shRNA 1. The original

pSilencer 2.1-U6 puro plasmid containing the scrambled shRNA was also transfected into the K562 cell to be used as a control.

3.7.1 Experimental design

Twelve 40 mL cultures of K562 (C1 to C12) cells were set up and incubated as indicated before until the cultures reached a cell density of 2.5 x 10^5 cells/mL. Then cultures C1 to C4 were transfected by electroporation with the pSilencer2.1-U6 puro plasmids containing the sense BGL3 shRNA 3, the sense BGL3 shRNA 4, the antisense BGL3 shRNA 1 and the scrambled shRNA respectively. Cultures C5 to C8 where transfected with the same plasmids as cultures C1 to C4 in order to obtain two biological samples (i.e. two independent K562 cell cultures) for each transfected plasmid. Cultures C9 to C12 followed the same electroporation procedure but without any plasmid in the electroporation cuvette (Figure 3-34).

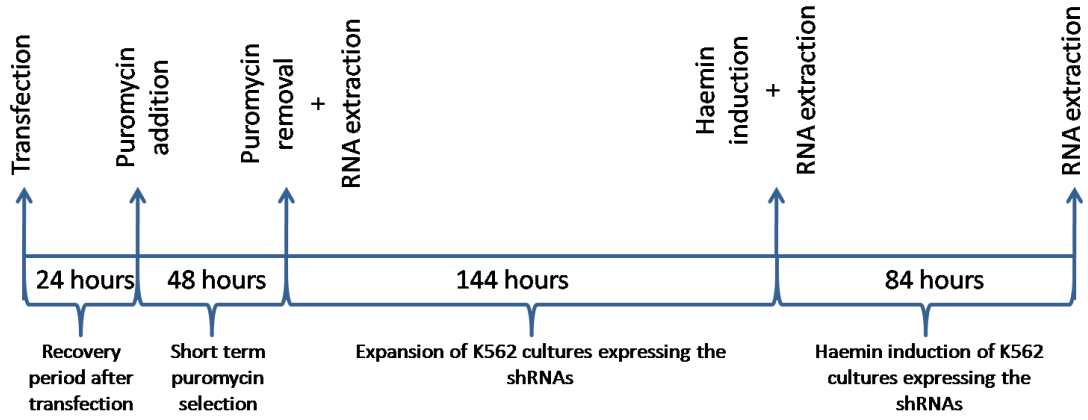


Figure 3- 34 Schematic overview of the experimental design after transfection of the K562 cells with the pSilencer constructs.

After transfection of the plasmids by electroporation into the K562 cells, the cells were left for recovery for 24 hours in normal culture conditions. A short term puromycin selection of 48 hours was used to enrich the cultures of cells that transiently expressed the shRNA insert (i.e. the cells that contained the plasmid). Thus, puromycin was added to the culture medium of cultures C1 to C8 to a final concentration of 1 µg/mL. According to Figure 3-32, after this 48 hours period of puromycin selection, it was expected to have killed approximately 80% of the K562 cells that were not transfected

with the ligated pSilencer2.1-U6 puro plasmids. This enrichment for transfected cells is useful for reducing background when analyzing BGL3 gene knockdown. C9 and C10 non-transfected control cultures were also subjected to puromycin selection as a control for the fraction of cells that survive selection. C11 and C12 non-transfected control cultures were grown without puromycin selection as a positive control for cell viability. The two cultures transfected with the plasmid encoding the scrambled hairpin siRNA, whose sequence has not homology with any region of the human genome, act as negative controls for analysis of BGL3 gene knockdown.

After 48 hours of puromycin selection, the cultures were pelleted and the medium containing puromycin was replaced by normal culture medium. Then cells were cultured in normal culture conditions for 144 hours to expand the cultures expressing the shRNAs. After this second recovery period, each culture was split again into two subcultures. One of the subcultures was supplemented with haemin to a final concentration of 100 μM to trigger the erythroid differentiation programme in the K562 cells. The purpose of this experiment was to check the effect of the knockdown of the BGL3 transcript on the expression levels of the β- and γ-globin genes during erythropoietic differentiation of K562 cells. RNA samples from the induced cultures were extracted after 84 hours of haemin addition to the K562 cultures. According to Figure 3-10, after 84 hours of haemin induction, the maximum expression levels of the BGL3 and γ-globin genes overlap. Thus, this is an ideal time point to check if the knockdown of the BGL3 transcript has worked and what are the effects of its silencing on the γ- and β-globin genes within the human β-globin locus during development.

3.7.2 β-globin locus expression profile in transiently transfected K562 cells

After transecton of K562 cells, the expression levels of the BGL3, γ- and β-globin genes were tested using quantitative RT-PCR (see section 2.3.17). The expression levels were normalized respect to the expression levels of endogenous GAPDH. Normalized fold expression of the genes mentioned above in transfected K562 cells was calculated respect to non-transfected control K562 cultures grown in equivalent conditions (i.e. cultures C9/C10). Results were analyzed using the software of the CFX96 Real-Time PCR Detection System. No genomic DNA was detected in the controls without reverese

transciptase. No PCR contamination was detected in the controls using water as a template.

A) Expression levels after 48 hours of puromycin selection

As shown in Figure 3-35, after 48 hours of puromycin selection the expression levels of the BGL3, β- and γ-globin genes in the K562 cells transfected with the pSilencer construct containing the scrambled shRNA insert showed no significant variations respect to the non-transfected cells. This indicates that the introduction of the pSilencer construct into the K562 cells has no effects on the expression levels of the studied genes.

Respect to non-transfected K562 cells, no significant variations were observed in the expression levels of the BGL3 transcript in the K562 cell cultures transfected with the pSilencer constructs containing the BGL3 sense shRNA 3 and BGL3 antisense shRNA 1. However, a significant decrease was observed in the expression levels of the BGL3 transcript in the K562 cell cultures transfected with the pSilencer construct containing the BGL3 sense shRNA 4. This indicates that this shRNA is effectively targeting the BGL3 transcript and triggering the RNAi pathway against it.

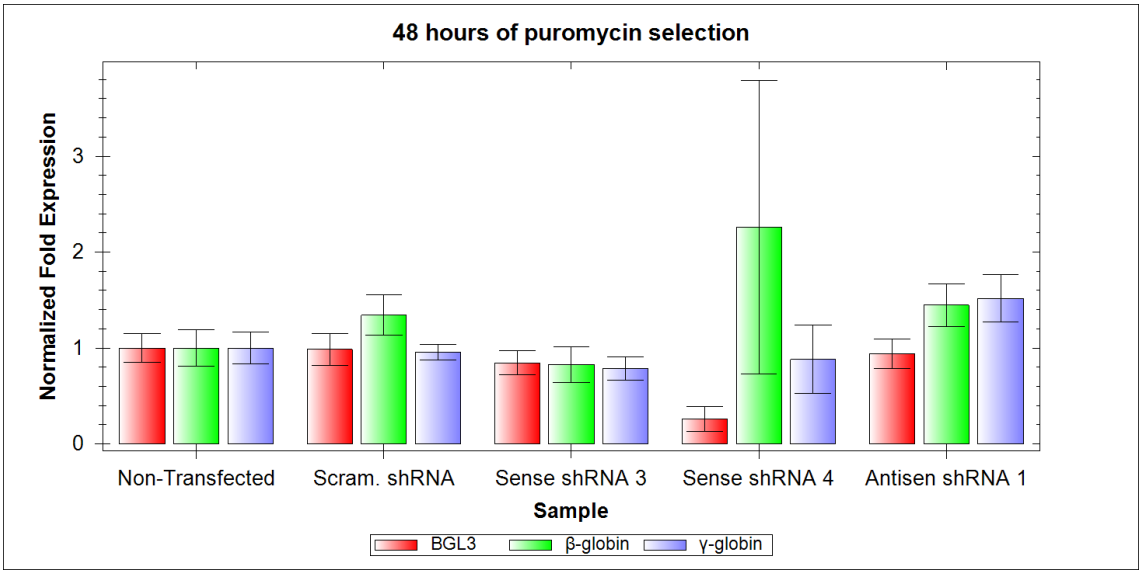


Figure 3- 35 β-globin locus gene expression profile of K562 transfected with the psilencer constructs containing the scrambled shRNA, the BGL3 sense shRNA 3, the BGL3 sense shRNA 4 and BGL3 antisense shRNA 1 inserts after 48 hours of puromycin selection.

Data obtained from three technical replicates of two biological replicates after 48 hours of puromycin selection of K562 cell cultures. Expression levels were normalized to endogenous GAPDH. Error bars represent standard error of the mean.

In the K562 cell cultures expressing the BGL3 sense shRNA 4 no significant variances are observed in the expression levels of the β- and γ-globin genes respect to non-transfected cells. However, looking at the expression of these genes in the two biological replicates of the experiment separately (Figures 3-36 and 3-37), it can be observed that the silencing of BGL3 transcript has only been achieved in the biological replicate 1 of the experiment. In this biological replicate, when the BGL3 transcription is effectively repressed there is a significant increment in the expression level of the β-globin gene of up to 3.5 fold. Of course that these data is not statistically significant because it is based in just one biological replicate, but it is indicating that suppression of BGL3 transcription could trigger the u-pregulation of the β-globin gene in K562 cells, a cell line characterized for its tendency to produce embryonic and foetal, but not adult haemoglobin. This could suggest that the BGL3 plays an important role in the

regulation of the human β-globin locus in K562 cells. More replicates are needed to test the reproducibility of these results.

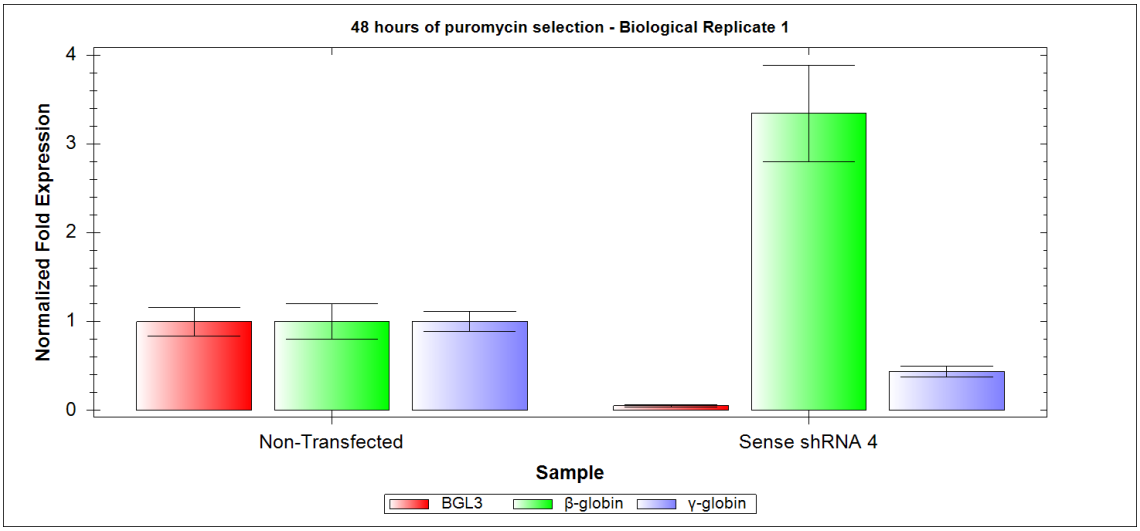


Figure 3- 36 Biological replicate 1 - β-globin locus gene expression profile of K562 transfected with the psilencer construct containing the BGL3 sense shRNA 4 insert after 48 hours of puromycin selection.

Data obtained from three technical replicates after 48 hours of puromycin selection of the K562 cell culture. Expression levels were normalized to endogenous GAPDH. Error bars represent standard error of the mean.

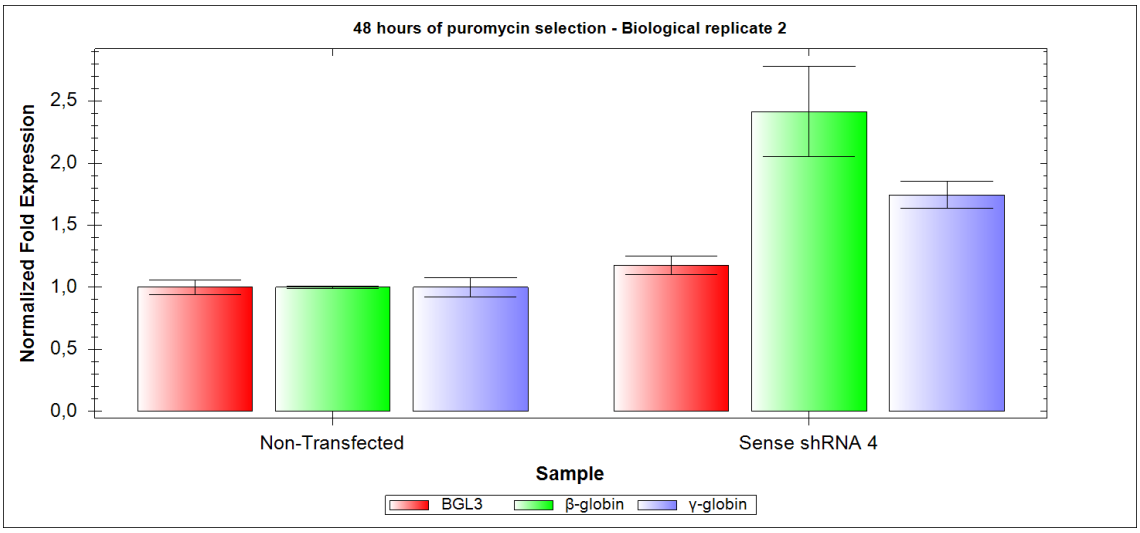


Figure 3- 37 Biological replicate 2 - β-globin locus gene expression profile of K562 transfected with the psilencer construct containing the BGL3 sense shRNA 4 insert after 48 hours of puromycin selection.

Data obtained from three technical replicates after 48 hours of puromycin selection of the K562 cell culture. Expression levels were normalized to endogenous GAPDH. Error bars represent standard error of the mean.

B) Expression levels after 144 hours of culture without puromycin

After 48 hours of puromycin selection, medium containing puromycin was replaced by normal culture medium. Then cells were cultured in normal culture conditions for 144 hours to expand the cultures expressing the shRNAs. After this period, RNA was extracted from all the transfected and non-transfected K562 cultures. Due to the financial constraints of this project, our attention was focused only in the cultures expressing the BGL3 sense shRNA 4, the only shRNA who proved to effectively target the BGL3 transcript. The RNA from these cultures was analyzed by quantitative RT-PCR using as controls the K562 cultures transfected with the pSilencer plasmid containing the scrambled shRNA insert and non-transfected K562 cultures (Figure 3-38).

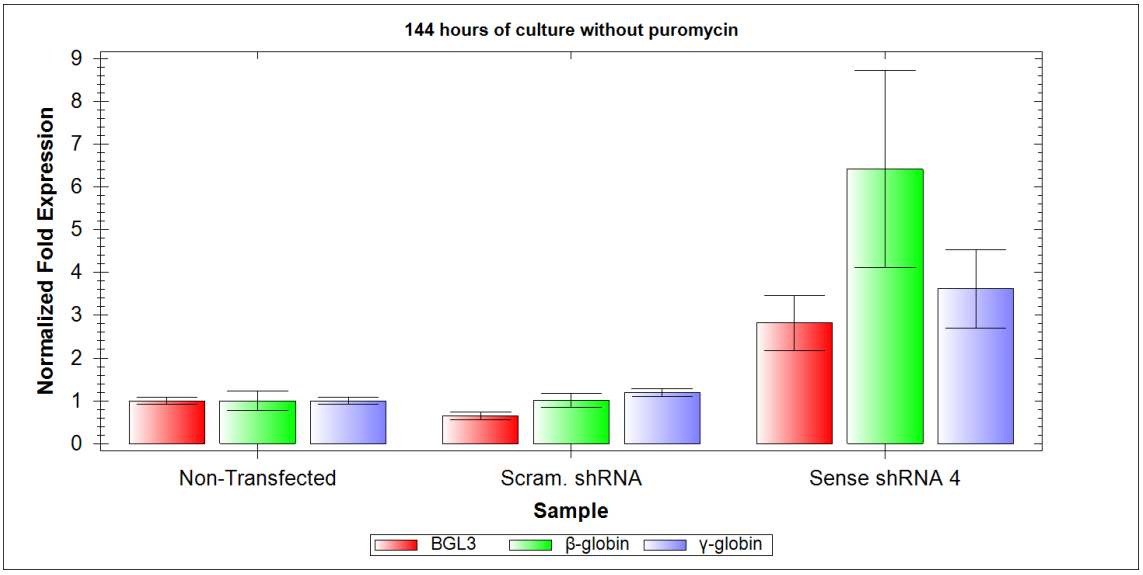


Figure 3- 38 β-globin locus gene expression profile of K562 cells transfected with the psilencer constructs containing the scrambled shRNA and the BGL3 sense shRNA 4 and inserts after 144 hours of culture without puromycin.

Data obtained from three technical replicates of two biological replicates after 144 hours of culture without puromycin. Expression levels were normalized to endogenous GAPDH. Error bars represent standard error of the mean.

As shown in Figure 3-38, there are no significant differences in the expression levels of the BGL3, β- and γ-globin genes between the non-transfected K562 cells and the K562 cells transfected with the construct containing the scrambled shRNA insert. However, significant differences can be observed between the non-transfected K562 cells and the K562 cells transfected with the construct containing the BGL3 sense shRNA 4 insert. The first surprise comes from the expression level of the BGL3 transcript, whose expression is three fold higher in the transfected cells. The expression levels of the β- and γ-globin genes are also higher (7 and 4 fold respectively) in the transfected cells.

C) Expression levels after 84 hours of culture in presence of haemin

After expanding the K562 cell cultures transiently expressing the shRNA 4 for 144 hours in normal condition cultures without puromycin, each culture was split again into two subcultures. One of the subcultures was supplemented with haemin to a final concentration of 100 µM to trigger the erythroid differentiation programme in the K562

cells. Induced K562 cells were cultured in presence for 84 hours, when it is expected that the maximum expression levels of the BGL3 and γ-globin genes overlap (Figure 3-10). Then RNA samples from the induced K562 cells transfected with the construct containing the scrambled shRNA insert and induced K562 cells expressing the BGL3 sense shRNA 4 were extracted to check if the knockdown of the BGL3 transcript was still working and what are the effects of its silencing on the γ- and β-globin genes within the human β-globin locus during development.

RNA samples from the induced K562 cell cultures expressing the BGL3 sense shRNA 4 were analyzed using as a control RNA samples from induced K562 cells transfected with the construct containing the scrambled shRNA insert. Since the transfection and haemin induction conditions were the same for both K562 cell cultures, any changes in the expression levels of the BGL3, β- and γ-globin genes can be attributed to the expression of the BGL3 sense shRNA 4.

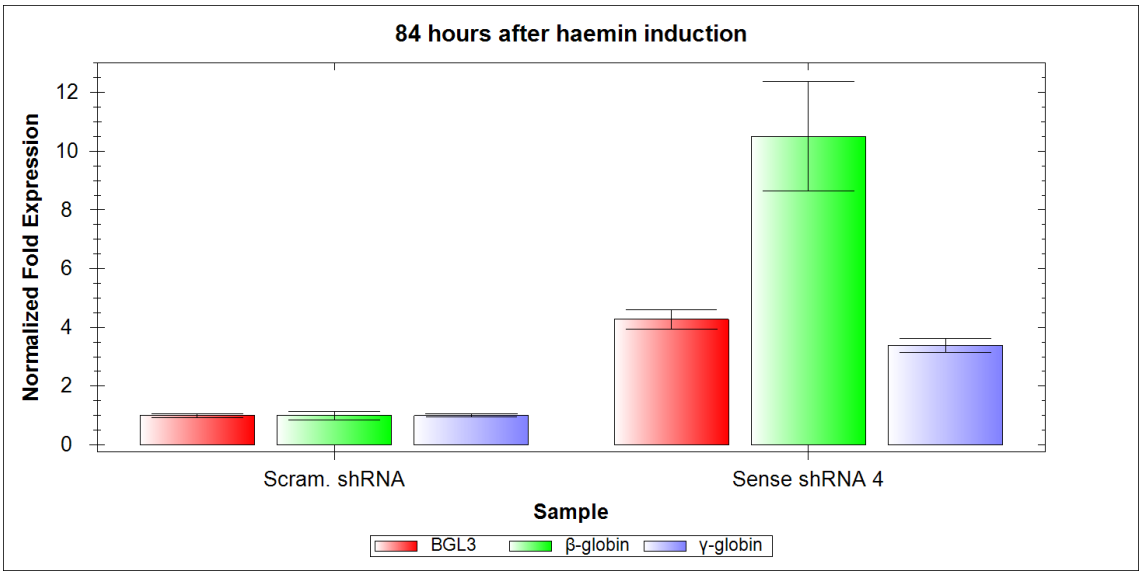


Figure 3- 39 β-globin locus gene expression profile of K562 cells transfected with the psilencer constructs containing the scrambled shRNA and the BGL3 sense shRNA 4 and inserts after 84 hours of haemin induction.

Data obtained from three technical replicates of two biological replicates after 84 hours of haemin induction. Expression levels were normalized to endogenous GAPDH. Error bars represent standard error of the mean.

As can be seen in Figure 3-39, the differences in the expression levels between the K562 cells transfected with the psilencer constructs containing the scrambled shRNA and the BGL3 sense shRNA 4 and inserts after 84 hours of haemin induction are similar to the observed in the last time point analysed (Figure 3-38). The expression levels of the genes studied, even the BGL3 transcript, are significantly increased in the cells expressing the BGL3 sense shRNA 4.

The only thing that really seems to be constant in all these experiments is that when the K562 cells express the BGL3 sense shRNA 4, which has been proved to effectively trigger the RNAi pathway against the BGL3 transcript, the levels of the β-globin gene are significantly increased. This fact is highly remarkable because the K562 cells are characterized for producing, when induced, large amounts of embryonic and foetal, but no adult, haemoglobin. This suggests that the BGL3 transcript could be involved, somehow, in repressing β-globin gene in K562 cells.

4 Discussion

4.1 Rationale of the project

Globin switching is a critical process for normal development, but it remains to be properly understood. Chromatin seems to play a role, but how the epigenetic status and structure of chromatin is regulated is not known. Recent evidence has emerged that non-coding RNAs play a role in regulating many aspects of cellular function, including chromatin dynamics. Here the role of a specific non-coding RNA, called BGL3, in regulating globin switching, was investigated.

4.2 Is BGL3 really a non-coding RNA?

As can be seen in Figures 3-2 and 3-3, the protein sequences for the three reading frames of both sense and antisense BGL3 transcripts contain high numbers of translation stop codons. This proves the low coding potential of the BGL3 transcript and strongly supports its status of non-coding RNA. According to its annotated length (1018 bp), the BGL3 transcript can be classified as a long non-coding RNA, a group of RNAs generally involved in gene silencing, DNA imprinting and demethylation, and in the generation of other RNA classes (Figure 1-10).

The nuclear distribution of the BGL3 transcript observed in FISH experiments also argues against a conventional coding role (Figure 1-23). Normally, FISH signals for genes are punctuate, with a cloud of mRNA in the cytoplasm (if using an exon probe). Non-coding RNAs tend to have a more diffuse pattern of FISH signals. This is the case, for example, of the Xist transcript, which coats chromosome X during X inactivation[68]. Similarly other non-coding RNAs involved in epigenetic processes, such as Kcnq1Ot1 appear to have signals whose spread is between Xist and ascent coding transcripts. Thus, with the pattern of FISH signals of the BGL3 transcript (Figure 1-23) it is tempting to speculate it may play a role in regulating chromatin architecture.

Several pieces of evidence support that non-coding RNAs are versatile molecules that, due to their chemical properties, can form complex structures that allow them to perform roles that were thought to be exclusive of proteins[39]. The prediction of the

secondary structure of the sense and antisense BGL3 transcripts suggests a potentially higher order conformation of the transcript, but analysis of the 1 kb either side of the BGL3 transcript suggest that these regions can also form similar higher order structures (not shown). Indeed, it is likely that programmes such as RNAfold are capable of turning any genomic region into a higher order structure. Furthermore, given that the majority of the genome is transcribed it is hard to find a non-transcribed region to serve as a control for a non-structured RNA. Further bioinformatic analysis is necessary to determine if the structures are statistically significant. It is also possible that the regions next to BGL3 are transcribed, separately or as part of a contiguous transcript. Further analysis of the region is necessary to determine the nature of the BGL3 transcript. For example its true size could be determined by northern blotting.

4.3 K562 cells as an erythroid model

Using as a marker of induced cells the increased levels of γ-globin and haemoglobin production, it has been shown that a program of erythroid maturation can be induced in our K562 culture by supplementing the culture medium with haemin.

As shown in Figure 3-6, there is a significant increase in the production of haemoglobin (presumably embryonic and foetal, but not adult, haemoglobin) in haemin-induced compared to non-induced K562 cell cultures. As previously show by other groups, induction of K562 cells with haemin results in an up-regulation of γ-globin but not significant increase in β-globin (Figure3-10). These results confirm the suitability of the K562 cell line as an erythroid model to investigate intergenic transcription within the β-globin locus.

Our results also show that the BGL3 transcript is specifically up-regulated during erythroid differentiation in K562 cells. Interestingly the peak of enrichment is 24 hours before that of γ-globin (Figure 3-10). This suggests that the BGL3 transcript could be is involved in activation of γ-globin gene. However, it is also plausible that the BGL3 could be involved in the repression of β-globin, or in a more complex mechanism involving the regulation of both γ- and β-globin genes. In any case, the induction experiments have to be repeated again in order to have more biological replicates to statistically validate the results presented here.

RNA FISH data from transgenic mice carrying the entire β-globin locus on a yeast artificial chromosome shows that the BGL3 transcript is expressed in embryonic and early foetal liver cells when the γ-domain of the β-globin locus is open, but later, when γ-globin gene expression shuts down, BGL3 transcription disappears (Unpublished work from the Fraser lab). This again suggests a role for the BGL3 transcript in regulating globin switching.

4.4 Generation of plasmids to perturb BGL3 expression

Whilst the results above suggest a regulatory role of the BGL3 transcript within the β-globin locus, the results are merely correlative; it is equally possible that the BGL3 transcript is switched on as a result of chromatin opening in the γ-domain of the locus. So it may be just a coincidence that BGL3 is expressed at the same time as γ-globin gene. To test this formally it is necessary to perturb the levels of the BGL3 transcript and test if this affects the expression levels of the genes of the β-globin locus.

4.4.1 pEF6/V5-His A constructs

In this project it was attempted to over-express the BGL3 transcript by cloning it into the constitutively expressed pEF6/V5-His A plasmid in order to study if the over-expression of the non-coding transcript perturbs the expression levels of the β- and γ-globin genes within the human β-globin locus in K562 cells. However, no *E. coli* clones carrying the pEF6/V5-His A plasmid containing the BGL3 inserts were isolated.

As shown in Figures 3-30 and 3-31, the isolated clones had no insert. The most likely explanation is that the original digested plasmid when purified included some non-digested plasmid. Then if the ligation was not very efficient (for various possible reasons, e.g. the PCR products were not efficiently digested, etc), the only circular plasmid introduced in *E. coli* was the original undigested plasmid. When transforming *E. coli*, the minus-insert control reaction yielded few colonies. This suggests that plasmids were able to re-circularise because of the inefficient purification of the digested plasmid. The failure to effectively remove the residual digest resulting from the digestion of the original pEF6/V5-His A plasmid would have lead to the re-ligation of the residual digest instead of the BGL3 sense and antisense inserts.

Primers only amplified the BGL3 sequence, yielding products of the expected size (Figure 3-24). No PCR contamination nor genomic DNA carry-over was detected in the control lanes of the electrophoresis of the PCR products. Plasmids digested with restriction enzyme yielded single bands equivalent to their molecular weight. However, some plasmid digests resulted in a series of faint bands as a result of the inefficiency of the restriction reactions (Figure 3-27). As shown in Figures 3-24 and 3-27, PCR products as well as linear pEF6/V5-His A plasmid were successfully extracted from agarose gels.

The most likely reason for the inefficiency of the ligation reactions is an incorrect molar ratio of plasmid to insert. Since spectrophotometric quantification of DNA is highly inaccurate, incorrect estimation of plasmid and insert concentrations could have prevented the optimal adjustment of the ligation reactions and, presumably, ligation reactions were set up at molar ratios which severely inhibited ligation of the insert to the plasmid. Thus, it is suggested to repeat the estimation of the purified linear plasmid and the digested PCR products to achieve the optimal 3:1 molar ratio using a more accurate spectrophotometer or non-spectrophotometric methods, such as estimation of band intensities on agarose gels. It would be worthwhile, as well, to set up parallel ligation reactions at a 1:1 insert to plasmid ratio in order to see if this is a more convenient ligation ratio for the specific case of the pEF6/V5-His A plasmid.

Due to the time limitations of this project, it was decided not to repeat all the cloning procedure for the pEF6/V5-His A plasmid and the BGL3 inserts. Attention was therefore focused on the pSilencer constructs.

4.4.2 pSilencer 2.1-U6 puro constructs

As mentioned before, it was decided to attempt to knockdown the BGL3 transcript using the pSilencer 2.1-U6 puro siRNA Expression Vector, which contains a U6 pol III promoter to express siRNAs and permits selection through an antibiotic resistance gene. The labour and cost of preparing bacterial cultures carrying the plasmids is significant. This was especially a problem in this project. A significant amount of time and work was spent setting up and optimizing an effective cloning procedure. It took several weeks to find the appropriate conditions to chemically transform our strain of K12 *E.*

coli. Furthermore, the fact of not working with the commercial kit of the pSilencer 2.1-U6 puro siRNA Expression Vector added more complexity to the project. Instead of the commercial kit, it was used a stock of pSilencer plasmid carrying the scrambled control. For this reason it was necessay to design a strategy to remove the original insert and purify the plasmid. This provided us with a relatively cheap tool to knockdown virtually any gene of the mammalian genome.

Once the protocols for the bacterial transformation and the ligation reactions were optimized, it was attempted to clone the four BGL3 sense shRNAs and the two BGL3 antisense shRNAs into the pSilencer 2.1-U6 puro siRNA Expression Vector. No remarkable complications were noticed during this stage of the project. Ligated plasmids were propagated, extracted and sent for sequencing to confirm that the inserts were correctly ligated into the plasmids and that there were no unwanted mutations. Sequencing reactions confirmed the presence of the correct shRNA inserts in the clone A of the sense BGL3 shRNA 3, clone C of the sense BGL3 shRNA 4 and clone A of the antisense BGL3 shRNA 1. Sequencing reactions for the clones containing the rests of shRNAs generated invalid short reads.

4.5 Transient transfection of K562 cells with ligated pSilencer 2.1-U6 puro plasmids

In order to perturb the expression levels of the BGL3 transcript by knocking it down using the RNAi pathway, K562 cells were transfected with the confirmed pSilencer constructs containing the inserts against the BGL3 transcript. Due to the time constraints of this project it was only attempted to transiently transfect K562 cells with the pSilencer constructs, instead of creating clonal cultures that stably express expresses the siRNA template introduced with the pSilencer vector. The electroporation method was used to transfect the K562 cells because is the procedure generally used for transfecting this cell line and a large number of protocols are available. However, other transfection methods for mammalian cells can be used, such as the calcium phosphate transfection method or cationic lipid-mediated transfection. The effect of the knockdown of the BGL3 transcript was tested on the activity of the γ- and β-globin genes, which were quantified using qRT-PCR, as well as the actual BGL3 transcript.

After 48 hours of puromycin selection for transfected cells, the expression level of the BGL3 transcript significantly decreased in the K562 cells transfected with the pSilencer construct containing the BGL3 sense shRNA 4 (Figure 3-35). This indicates that this shRNA effectively targets the BGL3 transcript and triggering the RNAi pathway against it.

Remarkably, when the BGL3 is effectively knocked down (only in one of the biological replicates; Figure 3-36), a significant increment in the expression level of the β-globin gene of up to 3.5 fold is observed. Always bearing in mind the poor statistical validity of these results, it can be suggested that the BGL3 transcript might be involved in the regulation of the β-globin locus, specifically in the inactivation of the β-globin gene during erythroid differentiation in K562 cells. Further work needs to be done in order to test the reproducibility of these preliminary results.

After 48 hours of puromycin selection, the transfected K562 cell cultures were cultured without puromycin for 144 hour to expand the cultures expressing the shRNAs. Analysis of the RNA of the samples from this time point showed that the expression level of the BGL3 transcript was three fold higher in the BGL3 sense shRNA 4-transfected cells. The expression levels of the β- and γ-globin genes were also higher (7 and 4 fold respectively) in these cells respect to non-transfected cells (Figure 3-38).

The same phenomenon was observed after 84 hours of haemin-induction in the comparison of the expression levels of the studied genes between the BGL3 sense shRNA 4-transfected K562 cells and K562 cells transfected with the construct containing the scrambled shRNA insert (Figure 3-39).

A possible explanation for these observations is that, since the K562 cultures were not subjected to a selection of stable transfected cell lines, cells could have stopped expressing the BGL3 sense shRNA 4. It is also possible that transfected cells surviving puromycin selection may not show a reduction in BGL3 transcription because may have found a way to mitigate its reduction by compensating it in another fashion or by shutting down expression of the siRNA. Another explanation could be that the siRNA against the BGL3 transcript also affects the expression level of GAPDH. This could cause inappropriate normalization of the expression levels of the studied genes in the samples expressing the BGL3 sense shRNA 4 and, therefore, invalidates the gene

expression data. This latter explanation is supported by the fact that the expression levels of GAPDH in the shRNA 4-transfected cultures are markedly lower than in the scrambled shRNA-transfected and non-transfected K562 cultures, as shown in Figures 4-1 and 4-2. As can be seen in these figures, looking at the relative fold expression of the studied genes without normalizing their expression to the endogenous GAPDH, it can be observed that the relative expression levels of GAPDH in the cultures transfected with the BGL3 sense shRNA 4 are markedly lower than in the transfection control cultures. However, the expression levels of the BGL3, β- and γ-globin genes are comparable between the cultures analyzed. It is also possible that GAPDH expression changes as erythropoiesis proceeds and that the cells are entering or leaving erythropoiesis as a result of BGL3 transcript knockdown by the BGL3 sense shRNA 4.

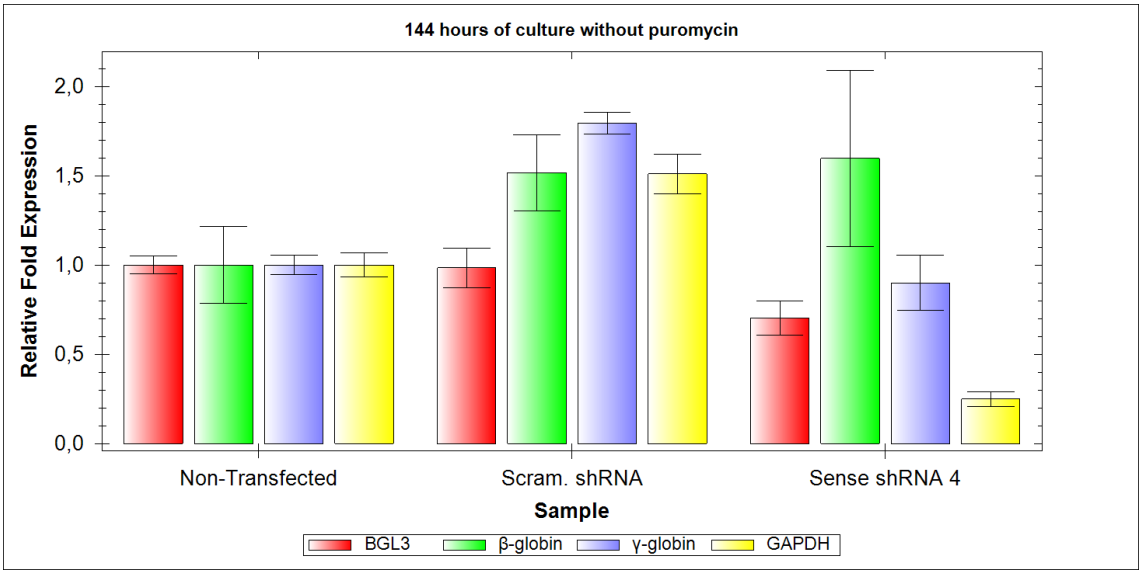


Figure 4- 1 β-globin locus gene expression profile of K562 cells transfected with the psilencer constructs containing the scrambled shRNA and the BGL3 sense shRNA 4 and inserts after 144 hours of culture without puromycin.

Data obtained from three technical replicates of two biological replicates after 144 hours of culture without puromycin. Error bars represent standard error of the mean.

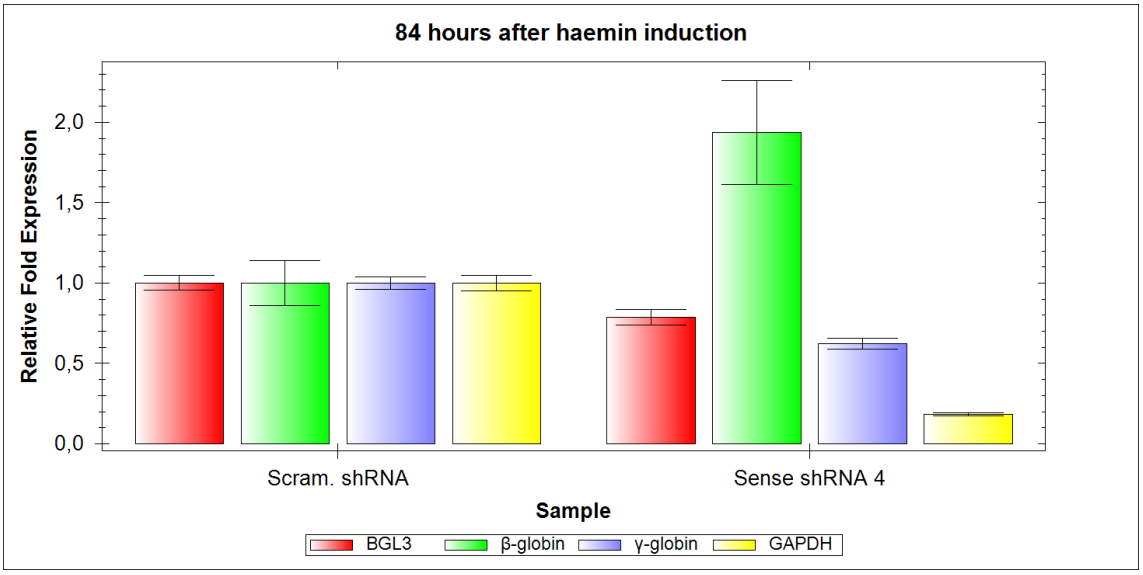


Figure 4- 2 β-globin locus gene expression profile of K562 cells transfected with the psilencer constructs containing the scrambled shRNA and the BGL3 sense shRNA 4 and inserts after 84 hours of haemin induction.

Data obtained from three technical replicates of two biological replicates after 84 hours of haemin induction. Error bars represent standard error of the mean.

An alternative hypothesis to explain the higher BGL3 levels observed in the shRNA 4 transfected cells respect to the control cultures could be that BGL3 may function generating siRNAs, since it has been shown that the BGL3 transcript is expressed in both sense and antisense directions (Figure 3-11). Thus, it is possible that the BGL3 transcript normally works by producing the two strands in opposite direction, which combine to form a dsRNA which is then processed in siRNAs by the RNAi pathway and is involved in regulating chromatin or transcripts within the β-globin locus. Then, knockdown of one strand could potentially prevent BGL3 from functioning correctly but without reducing the apparent level of BGL3 as detected by real time RT-PCR (because the method picks up both strands simultaneously). If one strand is removed by knockdown then the other strand is no longer being broken up by Dicer, so it begins to accumulate. This would explain the high levels of BGL3, as well as the variation in the expression of the γ- and β-globin genes, detected in the shRNA 4-transfected cells respect to the controls (Figures 3-38 and 3-39). This hypothesis could be tested by knocking down Dicer and measuring the effects on the expression levels of the BGL3.

4.6 Possible models for BGL3 action

Non-coding RNAs have been demonstrated to have key roles during cellular differentiation and organism development[38; 41; 42]. Non-coding transcription correlates with chromatin structural alterations and locus activation of various mammal loci, suggesting its involvement in the generation of open chromatin domains[50; 51].

The extent and regulatory roles of intergenic transcription throughout the β-globin locus remains largely unclear. One could assume that the function of intergenic transcription within the locus, in association with the β-LCR activity and histone modifications, is the generation and propagation of an open chromatin context in the locus to regulate transcription of its genes. However, non-coding throughout the locus could just be the consequence of its open conformation[24; 25].

The tracking model proposes that intergenic transcripts, together with the β-LCR, function in recruiting and delivering protein complexes and/or RNA polymerase II to the globin gene promoters[35]. The developmental regulation of intergenic transcription through the β-globin locus is supported by the observed correlation between its differential expression and the acetylation of the different subdomains of the locus at different developmental stages[7]. Since histone acetyltransferases and other chromatin-remodelling complexes such as SWI/SNF are associated with RNA–pol II complexes, intergenic transcription could play a role in decondensation of chromatin domains and gene activation[15].

There is a correlation between the differential expression of the genes of the β-globin locus during development and domain-wide epigenetic modifications in the locus[10]. The β-globin locus shows a complex pattern of histone modifications, different between foetal and adult erythrocytes, to epigenetically regulate the expression of the locus genes[50]. Our hypothesis is that non-coding RNAs coordinate these epigenetic marks within the β-globin locus, leading to altered chromatin dynamics and the subsequent accessibility of transcription factors to the enhancers and promoters of the different genes.

In light of our results, two explanations are possible to understand the mode of action of the BGL3 transcript and its effects on β-globin expression. The first explanation

considers that the BGL3 transcript functions in a similar way to the Air non-coding RNA (reviewed in section 1.2.3.1), which has been shown to interact with the H3K9 histone methyltransferase G9a[59].

Air is a clear example of a non-coding RNA that epigenetically silences transcription by recruiting to the promoter of the gene histone-modifying activities. The BGL3 transcript could be interacting with G9a as well, to epigentically silence expression of the β-globin gene (Figure 4-3)

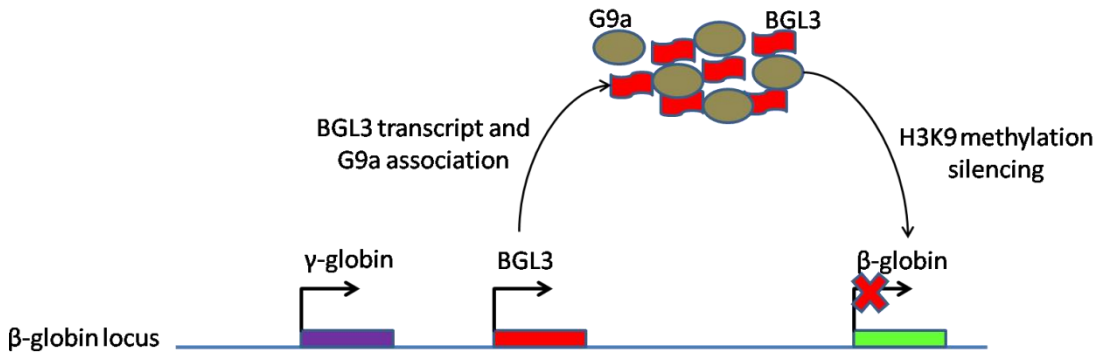


Figure 4- 3 Possible mode of action for the BGL3 transcript interacting with the H3K9 histone methyltransferase G9a.

In the proposed model shown in Figure 4-3, the BGL3 transcript associates with the H3K9 histone methyltransferase G9a to mediate H3K9 methylation silencing of the β-globin gene. This model might explain why it can be observed an increase in the expression level of the β-globin gene when effectively knockdown the BGL3 transcript is effectively knocked down via the RNAi pathway. However, this model cannot explain the increased level of BGL3 transcript observed after 144 hours of the transfection of K562 cells.

As mentioned before in this section, this increased levels of BGL3 transcript could be explained assuming that the transcript functions by forming a long double-stranded RNA. Recently, a new line of evidence supports that there is no correlation between intergenic transcription and chromatin activation in the β-globin locus. Instead, it is proposed that intergenic transcripts of the β-globin locus are specifically up-regulated in Dicer knockdown cells, involving the RNAi mechanisms in the regulation of intergenic transcription within the locus. Hence, intergenic transcription in the β-globin locus of

erythroid cells silences chromatin by default in regions not bound by activator transcriptional regulators[22]. According to this theory, the dsRNA formed by the sense and antisense BGL3 transcripts would be processed by Dicer into siRNAs that target the β-globin preventing in this way its expression (Figure 4-4).

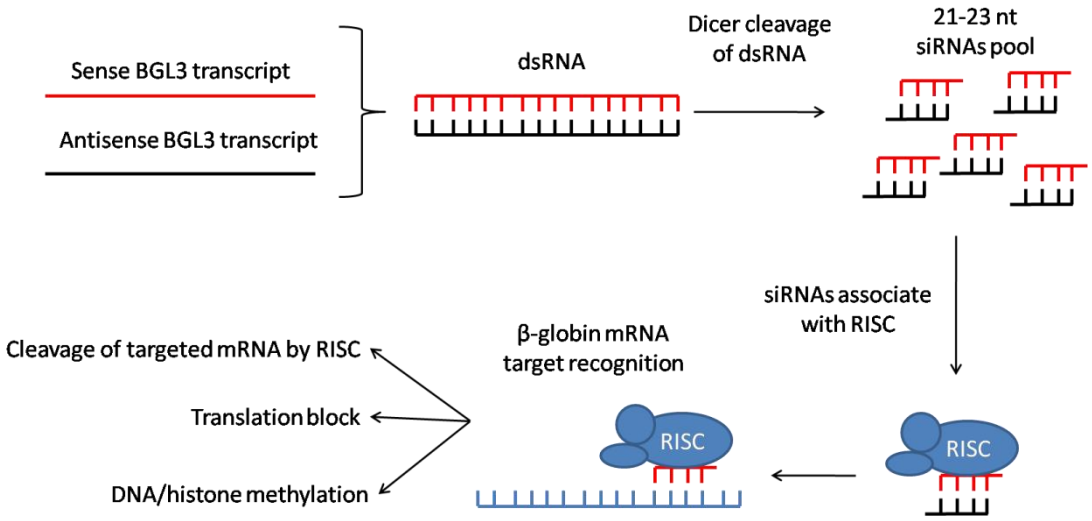


Figure 4- 4 Possible mode of action for the BGL3 transcript assuming that forms a dsRNA.

In Figure 4-4, the dsRNA formed by the sense and antisense BGL3 transcripts is processed into siRNAs, which are then loaded onto the RISC complex to finally target and cleave the β-globin mRNA. Alternatively, RISC could just block the transcription of the β-globin mRNA, or mediate epigenetic changes in its DNA sequence to epigenetically silence the β-globin gene. As indicated above, this model explains the increased level of BGL3 transcript after 144 hours the K562 cells transfection with the shRNAs against the non-coding transcript. Furthermore, this theory is supported by the fact that the BGL3 transcript seems to be produced in both sense and antisense directions (Figure 3-11).

Continued research is required in order to establish the influence of the BGL3 transcript upon the expression of γ- and β-globin genes and the rest of the genes of the β-globin locus. Further work needs to be done to explore the preliminary observations shown here. Generation of stable transfected K562 clones appears to be primordial to obtain robust and reproducible gene expression levels results. However, although the K562 cells have been proved to be a valid model to study intergenic transcription within the β-

globin locus, the biggest limitation of this project is the use of K562 cells. For example, the karyotype of the cells appears severely compromised (Figure 1-22). K562 cells carry three copies of the β-globin locus, but only two seem to be expressed[104]. This, added to the limitations always associated to the use of immortalized cell lines as models of study, suggests that the conclusions extracted from our studies on K562 cells could not be entirely applicable to normal human erythrocytes. Therefore it would be prudent to repeat the transfection of all the designed shRNAs against the BGL3 transcript, but this time using as a model for study *ex vivo* blood or transgenic mice containing the human β-globin locus. It is also important to study the effects of the over-expression of the BGL3 transcript. The efforts to clone the non-coding transcript into pEF6/V5-His A plasmid have to be repeated taking in account the suggestions indicated before. Over-expression of the BGL3 transcript may shed light on its role within the β-globin locus and on its involvement in the regulation of the γ- and β-globin genes.

5 Conclusions

This project aimed to better understand the regulation of the BGL3 non-coding transcript in the haemoglobin regulation process during erythropoiesis. Using quantitative RT-PCR, here it is shown that the BGL3 transcript is dynamically regulated during erythroid differentiation in haemin-induced K562 cells.

To address the potential function of the BGL3 transcript, it was knocked down using the RNAi pathway in the K562 cell line. Remarkably, expression of a siRNA hairpin construct directed against the BGL3 transcript led to transcriptional up-regulation of the β-globin gene in transiently transfected haemin-induced K562 cells, a cell line characterised for its capability to produce embryonic and foetal, but not adult, haemoglobin. Thus, our results suggest that BGL3 transcript knockdown may lead to the loss of the silencing of the adult haemoglobin gene in K562 cells.

Nevertheless, it is important to note that the findings described in this work are merely a starting point for the study of the regulatory role of the BGL3 transcript, and the results should be viewed with caution. More replicates are needed to prove the reproducibility of our results. Moreover, the transient transfections performed in this project only provide an initial view of the effects of knocking down the BGL3 transcript. Stable transfections will provide more accurate and reliable data about the impact that the BGL3 transcript silencing has on the expression levels of the γ- and β-globin genes.

Determining the exact functions of BGL3 transcript is beyond the scope of the present study. However, as indicated in the discussion and according to the literature available about the mode of action of other non-coding RNAs, our results suggest that BGL3 may mediate β-globin silencing via recruiting repressive histone-modifying activities to epigenetically silence transcription or via the RNAi pathway.

A plan of experiments similar to that described here, but using as a model of study *ex vivo* blood or humanised mice instead of the K562 cell line would provide more reliable data to elucidate the influence of intergenic transcription within the human β-globin locus has upon the developmentally regulated expression of the genes of the locus. This research is important as it is important to understand the regulation of γ-globin expression, which could be a target for the treatment of different blood disorders. These

types of disorders are amongst the most common genetic diseases in the world. Fully understanding of the regulatory mechanisms of intergenic transcription within the β-globin locus could provide us with the definitive knowledge to the eventual development of new therapies for β-thalassaemia or sickle cell disease based on the reactivation of γ-globin and HbF in adult erythroid cells.

In conclusion, our results are the first, to our knowledge, that describe a developmentally regulated expression of the BGL3 non-coding transcript in induced K562 cells, and provide evidence that suggests that this transcript may have a role in the silencing of the β-globin gene. Together, our results add to the growing body of evidence indicating that long non-coding transcripts are important molecules with key roles during development.

6 References

1. Hoffbrand, A. V., Pettit, J. E. and Moss, P. A. H. (2006), *Essential Haematology*, Fifth ed, Blackwell Publishing.

2. Hoffbrand, A. V., Catovsky, D. and Tuddenham, E. G. D. (2005), *Postgraduate Haematology*, Fifth ed, Blackwell Publishing.

3. Choong, M. L., Yang, H. H. and McNiece, I. (2007), "MicroRNA expression profiling during human cord blood-derived CD34 cell erythropoiesis", *Experimental Hematology*, vol. 35, no. 4, pp. 551-564.

4. Kanellopoulou, C. and Monticelli, S. (2008), "A role for microRNAs in the development of the immune system and in the pathogenesis of cancer", *Seminars in Cancer Biology*, vol. 18, no. 2, pp. 79-88.

5. Chen, C. -. and Lodish, H. F. (2005), "MicroRNAs as regulators of mammalian hematopoiesis", *Seminars in Immunology*, vol. 17, no. 2 SPEC. ISS., pp. 155-165.

6. Harju, S., McQueen, K. J. and Peterson, K. R. (2002), "Chromatin structure and control of β-like globin gene switching", *Experimental biology and medicine*, vol. 227, no. 9, pp. 683-700.

7. Orkin, S. H. (1995), "Regulation of globin gene expression in erythroid cells", *European Journal of Biochemistry*, vol. 231, no. 2, pp. 271-281.

8. Bank, A. (2006), "Regulation of human fetal hemoglobin: New players, new complexities", *Blood*, vol. 107, no. 2, pp. 435-443.

9. Liang, S., Moghimi, B., Yang, T. P., Strouboulis, J. and Bungert, J. (2008), "Locus control region mediated regulation of adult β-globin gene expression", *Journal of cellular biochemistry*, vol. 105, no. 1, pp. 9-16.

10. Stamatoyannopoulos, G. (2005), "Control of globin gene expression during development and erythroid differentiation", *Experimental hematology*, vol. 33, no. 3, pp. 259-271.

11. Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M. A., Sakaki, Y. and Sasaki, H. (2008), "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes", *Nature*, vol. 453, no. 7194, pp. 539-543.

12. Kiefer, C. M., Hou, C., Little, J. A. and Dean, A. (2008), "Epigenetics of β-globin gene regulation", *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 647, no. 1-2, pp. 68-76.

13. Fu, X. -., Liu, D. -. and Liang, C. -. (2002), "Chromatin structure and transcriptional regulation of the β-globin locus", *Experimental cell research*, vol. 278, no. 1, pp. 1-11.

14. Jackson, J. D., Miller, W. and Hardison, R. C. (1996), "Sequences within and flanking hypersensitive sites 3 and 2 of the β-globin locus control region required for synergistic versus additive interaction with the ε-globin gene promoter", *Nucleic acids research*, vol. 24, no. 21, pp. 4327-4335.

15. Tang, Y., Liu, D. -. and Liang, C. -. (2002), "Further understanding of the β-globin locus regulation at the molecular level: Looping or linking models?", *Genes to Cells*, vol. 7, no. 9, pp. 889-900.

16. Plant, K. E., Routledge, S. J. E. and Proudfoot, N. J. (2001), "Intergenic transcription in the human β-globin gene cluster", *Molecular and cellular biology*, vol. 21, no. 19, pp. 6507-6514.

17. Calzolari, R., McMorrow, T., Yannoutsos, N., Langeveld, A. and Grosveld, F. (1999), "Deletion of a region that is a candidate for the difference between the deletion forms of hereditary persistence of fetal hemoglobin and δβ-thalassemia

affects β- but not γ-globin gene expression", *EMBO Journal,* vol. 18, no. 4, pp. 949-958.

18. Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L. and Townes, T. M. (1990), "Human γ- to β-globin gene switching in transgenic mice", *Genes and Development,* vol. 4, no. 3, pp. 380-389.

19. Bender, M. A., Bulger, M., Close, J. and Groudine, M. (2000), "β-globin Gene switching and DNase I sensitivity of the endogenous β-globin locus in mice do not require the locus control region", *Molecular cell,* vol. 5, no. 2, pp. 387-393.

20. Fraser, P. (2006), "Transcriptional control thrown for a loop", *Current Opinion in Genetics and Development,* vol. 16, no. 5, pp. 490-495.

21. Arnosti, D. N. and Kulkarni, M. M. (2005), "Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards?", *Journal of cellular biochemistry,* vol. 94, no. 5, pp. 890-898.

22. Haussecker, D. and Proudfoot, N. J. (2005), "Dicer-dependent turnover of intergenic transcripts from the human β-globin gene cluster", *Molecular and cellular biology,* vol. 25, no. 21, pp. 9724-9733.

23. Hu, X., Eszterhas, S., Pallazzi, N., Bouhassira, E. E., Fields, J., Tanabe, O., Gerber, S. A., Bulger, M., Engel, J. D., Groudine, M. and Fiering, S. (2007), "Transcriptional interference among the murine β-like globin genes", *Blood,* vol. 109, no. 5, pp. 2210-2216.

24. Ho, Y., Elefant, F., Liebhaber, S. A. and Cooke, N. E. (2006), "Locus Control Region Transcription Plays an Active Role in Long-Range Gene Activation", *Molecular cell,* vol. 23, no. 3, pp. 365-375.

25. Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N. J. (1997), "Intergenic transcription and transinduction of the human β-globin locus", *Genes and Development,* vol. 11, no. 19, pp. 2494-2509.

26. Stamatoyannopoulos, G. (1991), "Human hemoglobin switching", *Science*, vol. 252, no. 5004, pp. 383.

27. Mcmorrow, T., Van Den Wijngaard, A., Wollenschlaeger, A., Van De Corput, M., Monkhorst, K., Trimborn, T., Fraser, P., Van Lohuizen, M., Jenuwein, T., Djabali, M., Philipsen, S., Grosveld, F. and Milot, E. (2000), "Activation of the β globin locus by transcription factors and chromatin modifiers", *EMBO Journal*, vol. 19, no. 18, pp. 4986-4996.

28. Schübeler, D., Francastel, C., Cimbora, D. M., Reik, A., Martin, D. I. K. and Groudine, M. (2000), "Nuclear localization and histone acetylation: A pathway for chromatin opening and transcriptional activation of the human β-globin locus", *Genes and Development*, vol. 14, no. 8, pp. 940-950.

29. Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. and Groudine, M. (1998), "The locus control region is necessary for gene expression in the human β-globin locus but not the maintenance of an open chromatin structure in erythroid cells", *Molecular and cellular biology*, vol. 18, no. 10, pp. 5992-6000.

30. Sankaran, V. G., Xu, J., Ragoczy, T., Ippolito, G. C., Walkley, C. R., Maika, S. D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M. A., Tucker, P. W. and Orkin, S. H. (2009), "Developmental and species-divergent globin switching are driven by BCL11A", *Nature*, vol. 460, no. 7259, pp. 1093-1097.

31. Enver, T., Raich, N., Ebens, A. J., Papayannopoulou, T., Constantini, F. and Stamatoyannopoulos, G. (1990), "Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice", *Nature*, vol. 344, no. 6264, pp. 309-313.

32. Raich, N., Enver, T., Nakamoto, B., Josephson, B., Papayannopoulou, T. and Stamatoyannopoulos, G. (1990), "Autonomous developmental control of human embryonic globin gene switching in transgenic mice", *Science*, vol. 250, no. 4984, pp. 1147-1149.

33. Choi, O. -. B. and Engel, J. D. (1988), "Developmental regulation of β-globin gene switching", *Cell*, vol. 55, no. 1, pp. 17-26.

34. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. -. and Fraser, P. (2002), "Long-range chromatin regulatory interactions in vivo", *Nature genetics*, vol. 32, no. 4, pp. 623-626.

35. Tuan, D., Kong, S. and Hu, K. (1992), "Transcription of the hypersensitive site HS2 enhancer in erythroid cells", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 23, pp. 11219-11223.

36. Quek, L. and Thein, S. L. (2007), "Molecular therapies in β-thalassaemia", *British journal of haematology*, vol. 136, no. 3, pp. 353-365.

37. Mercer, T. R., Dinger, M. E. and Mattick, J. S. (2009), "Long non-coding RNAs: insights into functions", *Nature Reviews Genetics*, .

38. Costa, F. F. (2007), "Non-coding RNAs: lost in translation?", *Gene*, vol. 386, no. 1-2, pp. 1-10.

39. Costa, F. F. (2005), "Non-coding RNAs: New players in eukaryotic biology", *Gene*, vol. 357, no. 2, pp. 83-94.

40. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E. and Chang, H. Y. (2007), "Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs", *Cell*, vol. 129, no. 7, pp. 1311-1323.

41. Mayer, C., Schmitz, K. -., Li, J., Grummt, I. and Santoro, R. (2006), "Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes", *Molecular Cell*, vol. 22, no. 3, pp. 351-361.

42. Bernstein, E. and Allis, C. D. (2005), "RNA meets chromatin", *Genes and Development*, vol. 19, no. 14, pp. 1635-1655.

43. Amaral, P. P. and Mattick, J. S. (2008), "Noncoding RNA in development", *Mammalian Genome,* vol. 19, no. 7-8, pp. 454-492.

44. Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G. and Orlando, V. (2007), "Noncoding RNA synthesis and loss of polycomb group repression accompanies the colinear activation of the human HOXA cluster", *RNA,* vol. 13, no. 2, pp. 223-239.

45. David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W. and Steinmetz, L. M. (2006), "A high-resolution map of transcription in the yeast genome", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, no. 14, pp. 5320-5325.

46. Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J. and Bähler, J. (2008), "Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution", *Nature,* vol. 453, no. 7199, pp. 1239-1243.

47. Mattick, J. S. (2004), "RNA regulation: A new genetics?", *Nature Reviews Genetics,* vol. 5, no. 4, pp. 316-323.

48. Krangel, M. S. (2007), "T cell development: Better living through chromatin", *Nature Immunology,* vol. 8, no. 7, pp. 687-694.

49. Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L. and Lander, E. S. (2006), "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells", *Cell,* vol. 125, no. 2, pp. 315-326.

50. Miles, J., Mitchell, J. A., Chakalova, L., Goyenechea, B., Osborne, C. S., O'Neill, L., Tanimoto, K., Engel, J. D. and Fraser, P. (2007), "Intergenic transcription, cell-cycle and the developmentally regulated epigenetic profile of the human beta-globin locus", *PLoS ONE,* vol. 2, no. 7.

51. Abarrategui, I. and Krangel, M. S. (2007), "Noncoding transcription controls downstream promoters to regulate T-cell receptor α recombination", *EMBO Journal,* vol. 26, no. 20, pp. 4380-4390.

52. Uhler, J. P., Hertel, C. and Svejstrup, J. Q. (2007), "A role for noncoding transcription in activation of the yeast PHO5 gene", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 19, pp. 8011-8016.

53. Schoenfelder, S., Smits, G., Fraser, P., Reik, W. and Paro, R. (2007), "Non-coding transcripts in the H19 imprinting control region mediate gene silencing in transgenic Drosophila", *EMBO Reports,* vol. 8, no. 11, pp. 1068-1073.

54. Shin, J. -., Fitzpatrick, G. V. and Higgins, M. J. (2008), "Two distinct mechanisms of silencing by the KvDMR1 imprinting control region", *EMBO Journal,* vol. 27, no. 1, pp. 168-178.

55. Spahn, L. and Barlow, D. P. (2003), "An ICE pattern crystallizes", *Nature Genetics,* vol. 35, no. 1, pp. 11-12.

56. Verona, R.I., Mann, M.R.W. and Bartolomei, M.S., (2003), *Genomic Imprinting: Intricacies of Epigenetic Regulation in Clusters*.

57. Thorvaldsen, J. L., Duran, K. L. and Bartolomei, M. S. (1998), "Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2", *Genes and Development,* vol. 12, no. 23, pp. 3693-3702.

58. Lyle, R., Watanabe, D., Te Vruchte, D., Lerchner, W., Smrzka, O. W., Wutz, A., Schageman, J., Hahner, L., Davies, C. and Barlow, D. P. (2000), "The imprinted antisense RNA at the lgf2r locus overlaps but does not imprint Mas1", *Nature Genetics,* vol. 25, no. 1, pp. 19-21.

59. Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R. and Fraser, P. (2008), "The Air noncoding RNA epigenetically silences

transcription by targeting G9a to chromatin", *Science,* vol. 322, no. 5908, pp. 1717-1720.

60. Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F. and Barlow, D. P. (1997), "Imprinted expression of the Igf2r gene depends on an intronic CpG island", *Nature,* vol. 389, no. 6652, pp. 745-749.

61. Zwart, R., Sleutels, F., Wutz, A., Schinkel, A. H. and Barlow, D. P. (2001), "Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes", *Genes and Development,* vol. 15, no. 18, pp. 2361-2366.

62. Andersen, A. A. and Panning, B. (2003), "Epigenetic gene regulation by noncoding RNAs", *Current Opinion in Cell Biology,* vol. 15, no. 3, pp. 281-289.

63. Sleutels, F., Zwart, R. and Barlow, D. P. (2002), "The non-coding Air RNA is required for silencing autosomal imprinted genes", *Nature,* vol. 415, no. 6873, pp. 810-813.

64. Muers, M. (2008), "RNA World: Antisense transcripts get involved", *Nature Reviews Genetics,* vol. 9, no. 12, pp. 898.

65. Woo, C. J. and Kingston, R. E. (2007), "HOTAIR Lifts Noncoding RNAs to New Levels", *Cell,* vol. 129, no. 7, pp. 1257-1259.

66. Mainguy, G., Koster, J., Woltering, J., Jansen, H. and Durtson, A. (2007), "Extensive polycistronism and antisense transcription in the mammalian Hox cluster", *PLoS ONE,* vol. 2, no. 4.

67. Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas III, E. J., Gingeras, T. R., Schreiber, S. L. and Lander, E. S. (2005), "Genomic maps and comparative analysis of histone modifications in human and mouse", *Cell,* vol. 120, no. 2, pp. 169-181.

68. Deng, X. and Meller, V. H. (2006), "Non-coding RNA in fly dosage compensation", *Trends in Biochemical Sciences,* vol. 31, no. 9, pp. 526-532.

69. Martens, J. A., Laprade, L. and Winston, F. (2004), "Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene", *Nature,* vol. 429, no. 6991, pp. 571-574.

70. Sonkoly, E., Bata-Csorgo, Z., Pivarcsi, A., Polyanka, H., Kenderessy-Szabo, A., Molnar, G., Szentpali, K., Bari, L., Megyeri, K., Mandi, Y., Dobozy, A., Kemeny, L. and Szell, M. (2005), "Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS", *Journal of Biological Chemistry,* vol. 280, no. 25, pp. 24159-24167.

71. Lu, W., Zhou, D., Glusman, G., Utleg, A. G., White, J. T., Nelson, P. S., Vasicek, T. J., Hood, L. and Lin, B. (2006), "KLK3IP is a novel androgen regulated and transcribed pseudogene of kallikreins that is expressed at lower levels in prostate cancer cells than in normal prostate cells", *Prostate,* vol. 66, no. 9, pp. 936-944.

72. Martens, J. A., Wu, P. -. J. and Winston, F. (2005), "Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae", *Genes and Development,* vol. 19, no. 22, pp. 2695-2704.

73. Johnson, K. D., Grass, J. A., Park, C., Im, H., Choi, K. and Bresnick, E. H. (2003), "Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain", *Molecular and cellular biology,* vol. 23, no. 18, pp. 6484-6493.

74. Ling, J., Baibakov, B., Pi, W., Emerson, B. M. and Tuan, D. (2005), "The HS2 enhancer of the β-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter", *Journal of Molecular Biology,* vol. 350, no. 5, pp. 883-896.

75. Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. and Groudine, M. (1998), "The locus control region is necessary for gene expression in the human β-globin

locus but not the maintenance of an open chromatin structure in erythroid cells", *Molecular and cellular biology*, vol. 18, no. 10, pp. 5992-6000.

76. Kim, A., Kiefer, C. M. and Dean, A. (2007), "Distinctive signatures of histone methylation in transcribed coding and noncoding human β-globin sequences", *Molecular and cellular biology*, vol. 27, no. 4, pp. 1271-1279.

77. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. and Fraser, P. (2000), "Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-globin locus", *Molecular cell*, vol. 5, no. 2, pp. 377-386.

78. Ristaldi, M. S., Drabek, D., Gribnau, J., Poddie, D., Yannoutsous, N., Cao, A., Grosveld, F. and Imam, A. M. A. (2001), "The role of the -50 region of the human γ-globin gene in switching", *EMBO Journal*, vol. 20, no. 18, pp. 5242-5249.

79. Applied Biosystems (2009), *'RNA interference research guide'*, available at: http://www.ambion.com/techlib/guides/Ambion_RNAi_res.pdf (accessed 11th June 2009).

80. Filipowicz, W. (2005), "RNAi: The nuts and bolts of the RISC machine", *Cell*, vol. 122, no. 1, pp. 17-20.

81. Anylam (2009), *'About RNAi'*, available at: http://www.alnylam.com/Leadership-in-RNAi/About-RNAi/index.php (accessed 11th June 2009).

82. Sontheimer, E. J. and Carthew, R. W. (2005), "Silence from within: Endogenous siRNAs and miRNAs", *Cell*, vol. 122, no. 1, pp. 9-12.

83. Downward, J. (2004), "Use of RNA interference libraries to investigate oncogenic signalling in mammalian cells", *Oncogene*, vol. 23, no. 51 REV. ISS. 8, pp. 8376-8383.

84. Sui, G., Soohoo, C., Affar, E. B., Gay, F., Shi, Y., Forrester, W. C. and Shi, Y. (2002), "A DNA vector-based RNAi technology to suppress gene expression in

mammalian cells", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 8, pp. 5515-5520.

85. Anylam (2009), *'RNA interference: An introduction to RNA interference (RNAi) and drug development'*, available at: http://www.alnylam.com/Files/Media_Kit/aln208_primer_07.pdf (accessed 11th June 2009).

86. Scherr, M. and Eder, M. (2007), "Gene silencing by small regulatory RNAs in mammalian cells", *Cell cycle (Georgetown, Tex.)*, vol. 6, no. 4, pp. 444-449.

87. Meister, G. and Tuschl, T. (2004), "Mechanisms of gene silencing by double-stranded RNA", *Nature*, vol. 431, no. 7006, pp. 343-349.

88. Applied Biosystems (2009), *'RNA Interference Overview'*, available at: http://ambion.com/techlib/resources/rnai/overview/index.html (accessed 11th June 2009).

89. Napoli, C., Lemieux, C. and Jorgensen, R. (1990), "Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans", *Plant Cell*, vol. 2, no. 4, pp. 279-289.

90. Zamore, P. D. and Haley, B. (2005), "Ribo-gnome: The big world of small RNAs", *Science*, vol. 309, no. 5740, pp. 1519-1524.

91. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998), "Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans", *Nature*, vol. 391, no. 6669, pp. 806-811.

92. Kim, V. N. (2003), "RNA Interference in Functional Genomics and Medicine", *Journal of Korean medical science*, vol. 18, no. 3, pp. 309-318.

93. Siomi, H. and Siomi, M. C. (2009), "On the road to reading the RNA-interference code", *Nature*, vol. 457, no. 7228, pp. 396-404.

94. Tomari, Y. and Zamore, P. D. (2005), "Perspective: Machines for RNAi", *Genes and Development,* vol. 19, no. 5, pp. 517-529.

95. Wadhwa, R., Kaul, S. C., Miyagishi, M. and Taira, K. (2004), "Know-how of RNA interference and its applications in research and therapy", *Mutation Research - Reviews in Mutation Research,* vol. 567, no. 1, pp. 71-84.

96. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001), "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells", *Nature,* vol. 411, no. 6836, pp. 494-498.

97. Lozzio, C. B. and Lozzio, B. B. (1975), "Human chronic myelogenous leukemia cell line with positive Philadelphia chromosome", *Blood,* vol. 45, no. 3, pp. 321-334.

98. Delgado-Cañedo, A., Santos, D. G. D., Chies, J. A. B., Kvitko, K. and Nardi, N. B. (2006), "Optimization of an electroporation protocol using the K562 cell line as a model: Role of cell cycle phase and cytoplasmic DNAses", *Cytotechnology,* vol. 51, no. 3, pp. 141-148.

99. Koeffler, H. P. and Golde, D. W. (1980), "Human myeloid leukemia cell lines: A review", *Blood,* vol. 56, no. 3, pp. 344-350.

100. Gribble, S. M., Roberts, I., Grace, C., Andrews, K. M., Green, A. R. and Nacheva, E. P. (2000), "Cytogenetics of the chronic myeloid leukemia-derived cell line K562 karyotype clarification by multicolor fluorescence in situ hybridization, comparative genomic hybridization, and locus-specific fluorescence in situ hybridization", *Cancer genetics and cytogenetics,* vol. 118, no. 1, pp. 1-8.

101. Rutherford, T. R., Clegg, J. B. and Weatherall, D. J. (1979), "K562 human leukaemic cells synthesise embryonic haemoglobin in response to haemin", *Nature,* vol. 280, no. 5718, pp. 164-165.

102. Fibach, E., Kollia, P., Schechter, A. N., Noguchi, C. T. and Rodgers, G. P. (1995), "Hemin-induced acceleration of hemoglobin production in immature cultured erythroid cells: Preferential enhancement of fetal hemoglobin", *Blood*, vol. 85, no. 10, pp. 2967-2974.

103. Ambion (2008), *'pSilencer puro Kit Protocol'*, available at: http://www.ambion.com/techlib/prot/fm_5762.pdf (accessed 11th June 2009).

104. Kim, A., Zhao, H., Ifrim, I. and Dean, A. (2007), "β-globin intergenic transcription and histone acetylation dependent on an enhance", *Molecular and cellular biology*, vol. 27, no. 8, pp. 2980-2986.